# Feedback from Automated Essay Evaluation Systems: A Review of Selected Research

**Yin Ling Cheung,** Nanyang Technological University, Singapore

## Introduction

In study of language assessment, particularly writing assessment, much attention has been paid to the potential of using automated scoring feedback systems to benefit learners (Carr, 2014; Chen & Cheng, 2008; Weigle, 2013a, 2013b). In 2013, *Assessing Writing* devoted an entire issue to investigating writing assessment with automated scoring systems (see Condon, 2013; Deane, 2013; Elliot & Williamson, 2013; Klobucar, Elliot, Deess, Rudniy, & Joshi, 2013; Ramineni, 2013; Ramineni & Williamson, 2013; Weigle, 2013b), calling for studies to address the impact of computer-generated feedback on the quality of writing. Likewise, in 2010, *Language Testing* launched a special issue on automated scoring and feedback systems (see Bernstein, Van Moere, & Cheng, 2010; Chapelle & Chung, 2010; Chodorow, Gamon, & Tetreault, 2010; Enright & Quinlan, 2010; Ginther, Dimova, & Yang, 2010; Weigle, 2010; Xi, 2010), featuring language testers' perception and engagement in automated feedback and scoring systems. To date, automated essay evaluation systems have received considerable attention from a number of research areas, ranging from rhetoric and composition studies, educational assessment, testing and measurement, to cognitive psychology and psycholinguistics (Elliot & Klobucar, 2013). In language testing, there is a large amount of research about the effectiveness of automated essay evaluation systems (Chodorow, et al., 2010; Hoang, 2011; Jones, 2006; Page, 2003; Warschauer & Ware, 2006), the effect of automated feedback on student writing (Elliot & Mikulas, 2004; Stevenson & Phakiti, 2014), and the perceptions of teachers and students on the use of automated essay evaluation systems (Klobucar, et al., 2013). This research resulted recently in a handbook of automated essay evaluation (Shermis & Burstein, 2013).

Papers in specialized journals in the field of applied linguistics, as well as those presented in conferences in the fields of language assessment and second/foreign language writing, have shown contrastive results in terms of the effectiveness

of automated essay evaluation systems. Up to now, the focus has mainly been on the accuracy, reliability, and meaning of automated feedback (Chodorow et al., 2010; Ferris, 2006), students' attitudes and teachers' attitudes (Chen & Cheng, 2008; Grimes & Warschauer, 2010), and impact of automated feedback on writing quality (El Ebyary & Windeatt, 2010; Rock, 2007; Schroeder, Grohe, & Pogue, 2008; Stevenson & Phakiti, 2014).

Generally, the literature (e.g., Elliot & Klobucar, 2013; Weigle, 2013a) suggests that automated essay evaluation systems are mainly used in making high-stakes decisions in large-scale tests, such as the Graduate Record Examination, the Test of English as a Foreign Language, the Graduate Management Admissions Test, the College Entrance Test in China (Zheng & Cheng, 2008), and the Entry Level Writing Requirement for all incoming freshmen enrolled in the University of California system (California, 2012), but their value as a tool for student learning is still inconclusive. On the one hand, scholars have pointed out that automated essay evaluation systems generally cannot judge the quality of argument and the intended meaning of a writer. Also, the systems can neither sort out all grammatical errors nor evaluate the authorial voice of a writer. The systems may identify certain errors wrongly. On the other hand, scholars (e.g., Weigle, 2013a) have emphasized the strengths of automated essay evaluation systems for identifying certain grammatical and mechanical errors, so that students will be able to make corrections before they submit the final versions of compositions. These systems may motivate students by providing them with immediate feedback on multiple drafts of their writing. In addition, automated feedback solves the problem of saving "face", which is difficult to avoid if reviews are carried out by peers or teachers. As pointed out by Dikli and Bleyle (2014) and Grimes & Warschauer (2010), there are differences between teacher feedback and automated feedback. In particular, automated feedback has an edge in giving quick feedback on multiple drafts, which may motivate students to revise their written work. Hence, automated feedback has the potential to act as a support tool that supplements teacher feedback.

However, regardless of scholars' views on the topic of automated feedback, we should recognize that the number of English Language teachers using automated scoring systems is growing rapidly. Since computers are programmed to provide scores and feedback on grammar and mechanics, teachers may reduce their time in grading surface level errors. Instead, they can focus on giving feed-

back with regard to higher-level issues of writing, such as authorial voice, development of ideas, and quality of arguments, which computers are not able to deal with at present. It is important to complement teachers' judgments of compositions with automated feedback because students benefit from the feedback from different perspectives. Since automated feedback has been considered controversial for learning purposes, compared with the traditional teacher feedback and peer feedback, there is a need to better understand the experience of those who use automated feedback in classrooms and what they think about the usefulness of computer-generated feedback.

Given the potential of using automated feedback to evaluate student writing performance, and the relative lack of knowledge of better motivating students in revising their essays, the review below reports on studies that focus on the use of automated essay evaluation systems as a tool for learning. The aim of the review is to highlight students' and teachers' perceptions of the pedagogical potential of automated essay evaluation systems. The review discusses main findings from the literature and implications of the research in second/foreign language education.

### Impact of Perceptions on Second Language Learning

The extent to which teachers' and students' perceptions influence the success of second language learning is an important area of investigation. Yet it has received little attention by researchers specializing in writing assessment. It is important to study perception because "students whose instructional expectations are not met may consciously or subconsciously question the credibility of the teacher and/or the instructional approach… Such lack of pedagogical face validity could affect learners' motivation" (Schulz, 1996, p. 349). Students' motivation to learn a second language could be enhanced if teachers can consider students' attitudes towards the instructional strategies (e.g., error correction strategies) that they adopt. Student attitudes have a mediating effect on written feedback (Sheen, 2011). This is to say, students having a goal to improve the accuracy of their writing tend to achieve higher levels of uptake (Hyland, 2003; Storch and Wigglesworth, 2010). Students who find that written feedback contradicts their beliefs in writing might choose not to respond to the feedback (Swain, 2006). Therefore, it is important to understand the role of perceptions toward feedback, including automated feedback practised in today's technology-assisted classrooms, in the learning process. This

understanding will help advance a dialog that elucidates the place of automated evaluation systems as a tool of teaching and learning writing.

**Teachers' and Students' Perceptions towards Automated Feedback as a Learning Tool**

Given increasing application of automated essay evaluation systems in classrooms, scholars outline the advantages of adopting them for learning and teaching (Attali & Burstein, 2006; Elliot, 2003; Landauer, Laham, & Foltz, 2003; Page, 2003; Weigle, 2013a). According to Weigle (2013a), software packages, particularly IntelliMetric and e-rater, can compute scores for writing faster than human raters, and can work on scores "with a high degree of accuracy" (p. 43), through an analysis of T-units and the length of certain discourse elements such as thesis statements, introduction, and conclusion. Another strength of IntelliMetric and e-rater is that they are unaffected by human factors such as distraction and fatigue, and correlate with the judgment of human raters in most of the cases in terms of grammar, mechanics, and organisation. These software tools are commonly used in judging argumentative, expository, and persuasive essay types.

How automated essay evaluation systems apply in placement, formative, and summative testing has been investigated by Condon (2013). Limitations of traditional grading by teachers, (i.e., time-consuming and labour intensive) may be partly overcome by the use of automated essay evaluation systems. It has been suggested that human raters should focus on the semantic level such as assessing development of ideas, argumentation, and presence of authorial voice, while allowing automated systems focus on mechanics and grammar. Nowadays, many students are computer literate. If they are willing to seek automated feedback from their school computer labs that have installed automated essay evaluation systems, they may get pointers to help revise their essays before final submission. In the long run, students may become more autonomous in their learning process.

However, there are challenges associated with feedback from automated essay evaluation systems. Weigle (2013a), Chodorow, et al. (2010) and Grimes and Warschauer (2010) list the performance characteristics of various automated feedback systems. E-rater and Criterion have relatively higher rates of identifying writing errors, compared with My Access. In particular, e-rater can be seen as an advanced evaluation tool because its accuracy rate is similar to human scoring

(Weigle, 2013a). Criterion has high rates in identifying errors, e.g., 80%. However, this system is weak in identifying errors related to articles and prepositions (Chodorow, et al., 2010). My Access has the lowest accuracy rate in pointing out any kinds of errors (Grimes & Warschauer, 2010). The performance characteristics of various systems suggest that the effectiveness of students' learning can be influenced by the software tool that their schools subscribe to. It could be problematic for students to depend solely on automated feedback because sometimes the software may falsely identify errors (Chodorow, Gamon, & Tetreault, 2010; Hoang, 2011; Weigle, 2013a). Also, the software cannot detect all mechanical and grammatical errors, such as subject-verb agreement and comma splices (Jones, 2006). Compared with computer evaluation systems, teachers are able to identify more writing errors beyond grammar, mechanics, and organization, and provide better quality feedback (Dikli & Bleyle, 2014). Better quality feedback refers to feedback that is accurate.

Some empirical research on automated feedback has focused on the perceptions of teachers and students towards its usefulness as a learning tool (Chen & Cheng, 2008; Grimes & Warschauer, 2006, 2008, 2010; Klobucar, et al., 2013; Link, Dursun, Karakaya, & Hegelheimer, 2014). A study conducted in secondary schools in the United States by Grimes and Warschauer (2010) highlighted teachers' strong disagreement with the scores generated by My Access, and the teachers' low confidence in the accuracy and reliability of automated feedback. Students in the study by Grimes and Warschauer (2010) also unfavorably judged the automated feedback they received, (though they might have been influenced by the teachers' negative attitude towards My Access). Chen & Cheng (2008) documented a similar perception when reporting that advanced Taiwanese EFL university students expressed that My Access was not useful to them. Particularly, the student participants pointed out that My Access could not provide them with concrete and explicit feedback on their essays in aspects such as content, coherence, and idea development. They believed that My Access could have been more useful to lower English language proficiency students acquiring basic writing skills.

From a different perspective, in Grimes and Warchauer's (2008) study, teachers listed a number of benefits related to the use of automated essay evaluation systems, ranging from saving teachers' time in giving feedback, enhancing students' motivation to revise their compositions before final submission, to promot-

ing student autonomy in learning. This positive perception is seen by another study (Grimes & Warchauer, 2006), whose results indicated that, as opposed to skepticisms that many teachers had for automated feedback, student participants in the study believed that when they saw an increase in the scores of their second and third drafts, they treated the increase in score as a sign of improvement in their writing.

Some points are worth mentioning to make the best out of automated essay evaluation systems. Automated feedback should not be used to replace teacher feedback (Baron, 1998; Chen & Cheng, 2008; Cheville, 2004; Conference on College Composition and Communication, 2004; Kellogg, Whiteford, & Quinlan, 2010; Phillips, 2007; Warschauer & Grimes, 2008). This is confirmed by Chen and Cheng's (2008) study that students would like computer-automated feedback to be followed by teacher feedback, and that teacher feedback is qualitatively and quantitatively not the same as computer automated feedback. Stevenson and Phakiti (2014) stressed that different software tools have different objectives. Hence, teachers and students should take note that the nature, quality, and effect of feedback provided by different software will be different.

## Research Questions

This brief overview of current literature on automated feedback, together with research findings on teachers' and students' perceptions of automated feedback as a tool for learning, provides us with a background to introduce the research issues. The purpose of this paper is to discuss the findings of a review of studies that can help shed light on teachers' and students' perceptions of utilizing automated feedback as a learning tool. The questions guiding this research are:

1. What are the teachers' perceptions towards the use of automated feedback systems in the writing classroom?
2. What are the students' perceptions towards the use of automated feedback systems in the writing classroom?

## Methodology

My intent was to review currently published refereed journal articles that discussed the impact of automated feedback on student writing, and students' and teachers' perceptions toward the use of automated feedback systems as a learning

tool. My primary source was international refereed journal articles whose authors employed rigorous methodologies. For the purposes of this study, conference papers, non-refereed articles, unpublished papers, master's theses, and doctoral dissertations were excluded as data sources. Recent papers, those published from 2008 through 2013, were selected for this review. For this study, I analyzed papers published in the following e-journals:

- Assessing Writing
- Journal of Technology, Learning and Assessment
- Language Learning and Technology
- Pedagogies: An International Journal

Most of the studies were conducted in the United States. One published study was conducted in an EFL setting in Taiwan.

## Findings and Discussion

This section highlights research findings from the review of the research studies. Findings are organized according to the two research questions. Specifically, teachers' perceptions of using an automated feedback system, and then students' perceptions.

### Teachers' Perceptions Toward the Use of Automated Feedback Systems

How teachers perceived the usefulness of automated feedback in the classroom was explored in four papers. Studies published in the past six years found contrasting results. Among these studies, Grimes and Warschauer (2008) conducted a mixed-method study that employed questionnaires, interviews, and class observations to study teachers who used automated feedback in four secondary schools in the United States. Most teachers expressed a positive attitude towards the use of automated feedback, as indicated in questionnaires and interviews, and confirmed by class observations. Specifically, the teachers saved time in grading the first drafts, as students were required to submit the earlier drafts to the automated evaluation system for feedback before the final submission to the teachers. Most teachers believed that, with automated essay evaluation, students could receive instant scores and specific feedback on grammar and mechanics. Based on the automated feedback, they could revise to improve their drafts in a more timely manner.

Grimes and Warschauer's (2008) study indicated teachers had favorable attitudes towards the use of automated feedback systems, whereas in another three studies (Cheng & Cheng, 2008; Grimes & Warschauer, 2010; Klobucar, et al., 2013) teachers showed negative attitudes toward the use of automated feedback. Chen and Cheng (2008) conducted their study through questionnaires, individual interviews with teachers, focus group interviews with the students, and student essays together with computer-generated feedback, in three classes of third-year undergraduate students in Taiwan. The results showed that all the three teachers had low confidence in the scores and feedback generated by the automated evaluation system. All the teachers supplemented the automated feedback with their own written feedback for all final drafts of the essays, and two of the teachers also provided feedback during students' revision process.

Two further papers report on teachers' negative perceptions of the use of automated feedback in classrooms. Grimes and Warschauer's (2010) study observes that teachers considered automated feedback to not be a useful tool for student learning. Their data consisted of surveys with teachers, interviews with teachers and administrators, class observations, student essays, and computer-generated feedback. Teachers complained about having to fix technical issues with the software, which took up precious classroom time. The teachers also exhibited low confidence in the scores generated by the automated evaluation system, similar to the findings of Chen and Cheng's (2008) study.

In another two-year study conducted by Klobucar et al. (2013) at a research university in the United States, results from the teacher survey highlighted that the teachers did not trust the accuracy of scores generated by the automated evaluation system. They believed that the automated feedback did not truly reflect the writing ability of the undergraduate students. The teachers indicated in the questionnaire that they had "limited desire to use the system in their composition classes" (p. 70).

The findings from the above-mentioned studies (Cheng & Cheng, 2008; Grimes & Warschauer, 2008, 2010; Klobucar, et al., 2013) suggest that automated evaluation can play a part in classroom, although teachers are not completely convinced of the accuracy of the computer-generated feedback. This statement is consistent with the recent research results reported by Elliot and Klobucar (2013) on automated essay evaluation and the teaching of writing. "Writing is a complex

socio-cognitive construct… automated essay evaluation does not read the essay but is trained to behave as a human rater would" (p. 20). Software, such as e-rater, may not be able to spot as many errors as teachers do (Weigle, 2013b). Hence, in a writing classroom, it will be useful if teachers can offer both automated feedback (e.g., surface level errors) and human rater feedback (e.g., global level issues such as flow of ideas, coherence, authorial voice, and argumentation) in the writing classroom. Also, teachers need to teach students explicitly how to interpret the automated scores and feedback. Teachers should remind the students to critically read and interpret automated feedback, and not to rely too much on the automated evaluation system because "One consistent concern, more specific to the computer, is that it cannot possibly use the same processes as humans in making discerning judgments about writing competence" (Shermis, Burstein, & Bursky, 2013, p. 2).

**Students' Perceptions Toward the Use of Automated Feedback Systems**

Of the studies examining students' perceptions towards the use of automated feedback systems, Grimes and Warschauer's (2008) study investigated automated writing assessment in four secondary classrooms in the United States. As indicated in the student surveys, about half of the students expressed that they were motivated to do more editing on surface level issues of their essays because of instant scores and immediate feedback provided by the software.

The majority of the studies, however, focused on reporting students' negative perceptions towards the use of automated feedback systems. Chen and Cheng's (2008) research described a group of third-year EFL undergraduates majoring in English and their experience using automated feedback system in the classroom. The study found that students reacted negatively towards the feedback system, because the feedback provided by the system was "abstract, vague, unspecific, formulaic, and repetitive" (p. 104). When asked what they considered to be the main problems with automated feedback, the majority of the students said that the automated feedback system "favored longer essays, overemphasized the use of transition words, discouraged creative ways of writing essays, and neglected global issues such as coherence and content development" (p. 104).

Grimes & Warschauer's (2010) study examined English-as-a-first-language students' attitude toward the use of automated feedback in facilitating the learning of writing. The study found that most of the students did not trust the scores pro-

vided by the evaluation system, as they believed that the score "did not reflect a human understanding of their essay" (p. 25). Many of them commented that they did not even bother to read pages of suggestions generated by the essay evaluation system, showing that they took the automated feedback lightly.

Klobucar et al. (2013) studied the implementation of automated essay scoring in identifying at-risk students in a freshmen writing course in the United States. The students commented that the automated evaluation system was unable to give accurate feedback that reflected their writing ability. In particular, "students reported that they would have done better had they written to humans instead of machines" (p. 70). They did not believe that it would be useful to integrate the automated feedback in the classroom. They also were skeptical about using the scores from the evaluation system in summative assessment, due to accuracy and reliability issues of the system.

Overall, students showed an unfavorable attitude towards the use of automated evaluation systems (as they now exist) in the classroom. As noted above, automated scoring systems, unlike human beings, are unable to judge certain important components in an essay, such as how well the essay addresses the topic, and other aspects including coherence, appropriate choice of word, and the development of ideas (Enright & Quinlan, 2010). The perceived preference for longer essays by automated evaluation systems was questioned by many students in Chen and Cheng's (2008) study and may be explained by the fact that automated essay scoring software seem to consistently over-value the number of words in calculating student writing scores (Beckman, 2010; Kobrin, Deng, & Shaw, 2007, 2011; Perelman, 2012, 2014; Winerip, 2012). It should be pointed out, however (as noted by Chandrasegaran, 2013) in human scoring, longer essays also tend to score higher. It is not just because they contain more words. Rather, it is that longer essays usually have more adequately developed arguments. It takes more words to develop sophisticated arguments, and perform maturity-projecting acts like hedging appropriately and showing awareness of audience. In summary, in this brief review of some recent studies, it is hoped that teachers' and students' perceptions toward automated feedback will advance a constructive dialog to facilitate a better understanding of the place of automated evaluation system in classrooms as a tool of learning writing.

## Conclusion

This study aimed to discuss findings in a review of studies that can help us understand teachers' and students' perception of utilizing automated feedback as a tool of learning. The findings from the review of studies (Cheng & Cheng, 2008; Grimes & Warschauer, 2008, 2010; Klobucar, et al., 2013) suggest that automated evaluation system plays a part in the classroom in motivating students to revise their drafts, but that the majority of teachers and students are still skeptical of the accuracy of the computer-generated feedback. The specific reactions could depend on the type of software selected for use in the classrooms. Second/foreign language writing teachers should consider adopting software demonstrated to have a higher accuracy rate, such as Criterion. Teachers would also need to play an active role in helping students interpret the scores and feedback provided by the software during the writing process. Since most of the published studies were conducted in English-as-a-first-language settings (i.e., the United States), whether there is a varying relationship between the effectiveness of software and the first language of students is a topic that merits further research. Findings from this review of studies indicate that few students and teachers exhibit a predominantly favorable view of using automated essay evaluation systems. More L2 writing specialists, in addition to software vendors and language assessment experts, should further examine how to design writing tasks to harness specific strengths of automated essay evaluation systems, as well as further refine the systems to handle effectively a wider variety of writing tasks.

Although caution is needed to generalize the findings based on automated evaluation systems from a small number of studies, we can draw two important implications pertaining to the impacts of perceptions on second language learning. Firstly, the above review has shown that students' attitude towards feedback (of any kind) on writing can be negative at times (Storch & Wigglesworth, 2010; Swain, 2006; Swain & Lapkin, 2002). For example, some students may resist written feedback due to a wrong impression that only unskilled writers need advice in revising their writing. Accepting feedback from a non-human source would only increase the resistance. Teachers might deal with the negative attitude by convincing their students that writing is inherently a recursive and complex activity. Novice and seasoned writers alike need to go through multiple rounds of revisions to improve the quality of their work, in which they may benefit tremendously from the

views and reactions of others. Secondly, to improve teachers' acceptance of the use of automated evaluation systems, Heads of Departments and academic writing specialists can act as facilitating agents of change. Teachers should be encouraged to participate in seminars on the use of automated evaluation systems in writing classrooms, and training workshops where they gain hands-on knowledge in using these systems to give constructive feedback. On the other hand, teachers need to know that automated evaluation systems ideally play an assistive role only; they should not replace the teachers themselves in the total learning process including personal feedback from the teachers.

**Limitations and Suggestions for Future Research**

It must be first said that this study was limited in its coverage of the major sources that may be relevant to the topic at hand. My first suggestion is for the carrying out of a much larger study, involving other major journals such as the *Journal of Writing Assessment* and the *Journal of Second Language Writing*, as well as a carefully-selected number of well-designed theses and dissertations. With that caveat in mind, this review study has pointed out some of the issues and the importance of attitudes and perceptions in second language learning. Future research might also examine the additional impacts of everyday practical challenges that students face, e.g., access, the technological learning curve, and the demands of time and persistence in revising familiar ideas and texts carefully and repetitively. Moreover, the usefulness and accuracy of automated evaluation systems on genres other than essays should be investigated. In light of the increasing role of automated evaluation learning systems as assistive tools, it would also be interesting to investigate the role of teachers in helping students interpret the feedback from these systems properly, or avoid pitfalls from careless or overly dependent use.

## References

Attali, Y., & Burstein, J. (2006). Automated essay scoring with e-rater V.2. *Journal of Technology, Learning, and Assessment, 4*(3), 3-30.

Baron, D. (1998, November 20). When professors get A's and the machines get F's. *Chronicle of Higher Education, A56.*

Beckman, M. (2010). Quality vs. quantity: How to score higher on the SAT essay component: Has teen unlocked secret to a better SAT score? *Good Morning America.* Retrieved from http://abcnews.go.com/GMA/ConsumerNews/teen-student-finds-longer-sat-essay-equals-score/story?id=12061494

Bernstein, J., Van Moere, A., & Cheng, J. (2010). Validating automated speaking tests. *Language Testing, 27*, 355-377.

California, Regents of the University of California. (2012). University of California Office of the President Student Affairs: Entry Level Writing Requirements, from http://www.ucop.edu/elwr/index.html

Carr, N. T. (2014). Computer-automated scoring of written responses. In Kunnan, A. (Ed.), *The companion to language assessment* (pp. 1063-1078). Hoboken, NJ: Wiley-Blackwell.

Chandrasegaran, A. (2013). The effect of a socio-cognitive approach to teaching writing on stance support moves and topicality in students' expository essays. *Linguistics and Education, 24*, 101-111.

Chapelle, C. A., & Chung, Y. (2010). The promise of NLP and speech processing technologies in language assessment. *Language Testing, 27*, 301-315.

Chen, C.-F., & Cheng, W.-Y. (2008). Beyond the design of automated writing evaluation: Pedagogical practices and perceived learning effectiveness in EFL writing classes. *Language Learning & Technology, 12*(2), 94-112.

Cheville, J. (2004). Automated scoring technologies and the rising influence of error. *English Journal, 93,* 47-52.

Chodorow, M., Gamon, M., & Tetreault, J. (2010). The utility of article and preposition error correction systems for English language learners: Feedback and assessment. *Language Testing, 27,* 419-436.

Condon, W. (2013). Large-scale assessment, locally-developed measures, and automated scoring of essays: Fishing for red herrings? *Assessing Writing, 18,* 100-108.

Conference on College Composition and Communication. (2004). *CCCC position statement on teaching, learning, and assessing writing in digital environments.* Retrieved from htttp://www.ncte.org/cccc/resources/positions/digital-environments

Deane, P. (2013). On the relation between automated essay scoring and modern views of the writing construct. *Assessing Writing, 18,* 7-24.

Dikli, S., & Bleyle, S. (2014). Automated essay scoring feedback for second language writers: How does it compare to instructor feedback? *Assessing Writing, 22,* 1-17.

El Ebyary, K., & Windeatt, S. (2010). The impact of computer-based feedback on students' written work. *International Journal of English Studies, 10*(2), 121-142.

Elliot, S. (2003). IntelliMetric: from here to validity. In M. D. Shermis & J. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp. 71-86). Mahwah, NJ: Erlbaum Associates.

Elliot, N., & Klobucar, A. (2013). Automated essay evaluation and the teaching of writing. In M. D. Shermis & J. Burstein (Eds.), *Handbook of automated essay scoring* (pp. 16-35). New York and London: Routledge.

Elliot, N., & Williamson, D. M. (2013). Assessing Writing special issue: Assessing writing with automated scoring systems. *Assessing Writing, 18*(1), 1-6.

Elliot, S. M., & Mikulas, C. (2004, April). *The impact of My Access use on student writing performance: A technology overview and four studies.* Paper presented at the Annual meeting of the American Educational Research Association, San Diego, California.

Enright, M., & Quinlan, M. (2010). Complementing human judgment of essays written by English language learners with e-rater scoring. *Language Testing, 27,* 317-334.

Ferris, D. (2006). Does error feedback help student writers? New evidence on the short- and long-term effects of written error correction. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 81-104). Cambridge, UK: Cambridge University Press.

Ginther, A., Dimova, S., Yang, R. (2010). Conceptual and empirical relationships between temporal measures of fluency and oral English proficiency with implications for automated scoring. *Language Testing, 27,* 379-399.

Grimes, D., & Warschauer, M. (2006, April). *Automated essay scoring in the classroom.* Paper presented at the American Educational Research Association Annual Conference, San Francisco, California.

Grimes, D., & Warschauer, M. (2008). Learning with laptops: A multi-method case study. *Journal of Educational Computing Research, 38*(3), 305-332.

Grimes, D., & Warschauer, M. (2010). Utility in a fallable tool: A multi-site case study of automated writing evaluation. *Journal of Technology, Learning, and Assessment, 8*(6), 1-43.

Hoang, G. (2011). *Validating My Access as an automated writing instructional tool for English language learners* (Unpublished master's thesis). California State University, Los Angeles.

Hyland, F. (2003). Focusing on form: Student engagement with teacher feedback. *System, 31*, 217-230.

Jones, E. (2006). ACCUPLACER's essay-scoring technology: When reliability does not equal validity. In P. F. Ericsson & R. H. Haswell (Eds.), *Machine scoring of student essays: Truth and consequences* (pp. 93-113). Logan: Utah State University Press.

Kellogg, R., Whiteford, A., & Quinlan, T. (2010). Does automated feedback help students learn to write? *Journal of Educational Computing Research, 42*, 173-196.

Klobucar, A., Elliot, N., Deess, P., Rudniy, O., & Joshi, K. (2013). Automated scoring in context: Rapid assessment for placed students. *Assessing Writing, 18*, 62-84.

Kobrin, J. L., Deng, H., & Shaw, E. J. (2007). Does quantity equal quality?: The relationship between length of response and scores on the SAT essay. *Journal of Applied Testing Technology, 8*(1), 1-15.

Kobrin, J. L., Deng, H., & Shaw, E. J. (2011). The association between SAT prompt characteristics, response features, and essay scores. *Assessing Writing, 16*, 154-169.

Landauer, T. K., Laham, D., & Foltz, P. W. (2003). Automated scoring and annotation of essays with the Intelligent Essay Assessor. In M. D. Shermis & J. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp. 87-112). Mahwah, NJ: Erlbaum Associates.

Link, S., Dursun, A., Karakaya, K., & Hegelheimer, V. (2014). Towards best ESL practices for implementing automated writing evaluation. *CALICO Journal, 31*, 323-344.

Page, E. B. (2003). Project Essay Grade: PEG. In M. D. Shermis & J. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp. 43-54). Mahwah, NJ: Lawrence Erlbaum.

Perelman, L. (2012). Length, score, time, and construct validity in holistically graded writing assessments: The case against automated essay scoring (AES). In C. Bazerman, C. Dean, K. Lunsford, S. Null, P. Rogers, & A. Stansell, et al. (Eds.), *New directions in international writing research* (pp. 121-132). Anderson, SC: Parlor Press.

Perelman, L. (2014). When "the state of the art" is counting words. *Assessing Writing, 21*, 104-111.

Phillips, S. (2007). *Automated essay scoring: A literature review*. Kelowna, B.C.: Society for the Advancement of Excellence in Education.

Ramineni, C. (2013). Validating automated essay scoring for online writing placement. *Assessing Writing, 18*, 40-61.

Ramineni, C., & Williamson, D. (2013). Automated essay scoring: Psychometric guidelines and practices. *Assessing Writing, 18*, 25-39.

Rock, J. (2007). *The impact of short-term use of Criterion on writing skills in ninth grade* (Research Report 07-07). Princeton, NJ: Educational Testing Service.

Schroeder, J., Grohe, B., & Pogne, R. (2008). The impact of criterion writing evaluation technology on criminal justice student writing skills. *Journal of Criminal Justice Education, 19*, 432-445.

Schulz, R. A. (1996). Focus on form in the foreign language classroom: Students' and teachers' views on error correction and the role of grammar. *Foreign Language Annals, 29*, 343-364.

Sheen, Y. (2011). *Corrective feedback, individual differences and second language learning*. New York: Springer.

Shermis, M. D., & Burstein, J. (Eds.). (2013). *Handbook of automated essay evaluation: Current applications and new directions*. New York, NY: Routledge.

Shermis, M. D., Burstein, J., & Bursky, S. A. (2013). Introduction to automated essay evaluation. In M. D. Shermis & J. Burstein (Eds.), *Handbook of automated essay scoring* (pp. 1-15). New York and London: Routledge.

Stevenson, M., & Phakiti, A. The effects of computer-generated feedback on the quality of writing. *Assessing Writing, 19*, 51-65.

Storch, N., & Wigglesworth, G. (2010). Learners' processing, uptake, and retention of corrective feedback on writing: Case studies. *Studies in Second Language Acquisition, 32*, 303-334.

Swain, M. (2006). Languaging, agency and collaboration in advanced language proficiency. In H. Byrnes (Ed.), *Advanced language learning: The contribution of Halliday and Vygotsky* (pp. 95-108). New York: Continuum.

Swain, M., & Lapkin, S. (2002). Talking it through: Two French immersion learners' response to reformulation. *International Journal of Educational Research, 37*, 285-304.

Warschauer, M ., & Grimes, D. (2008). Automated writing assessment in the classroom. *Pedagogies: An International Journal, 3*(1), 22-36.

Warschauer, M., & Ware, P. (2006). Automated writing evaluation: Defining the classroom research agenda. *Language Teaching Research, 10*(2), 1-24.

Weigle, S. C. (2010). Validation of automated scores of TOEFL iBT tasks against non-test indicators of writing ability. *Language Testing, 27*, 335-353.

Weigle, S. C. (2013a). English as a second language writing and automated essay evaluation. In M. D. Shermis & J. Burstein (Eds.), *Handbook of automated essay scoring* (pp. 36-54). New York and London: Routledge.

Weigle, S. C. (2013b). English language learners and automated scoring of essays: Critical considerations. *Assessing Writing, 18*, 85-99.

Winerip, M. (2012). Facing a Robo-Grader? Just keep obfuscating mellifluously. *The New York Times*. Retrieved from http://www.nytimes.com/2012/04/23/education/robo-readers-used-to-grade-test-essays.html

Xi, X. (2010). Automated scoring and feedback systems: Where are we and where are we heading? *Language Testing, 27*, 291-300.

Zheng, Y., & Cheng, L. (2008). College English Test (CET) in China. *Language Testing, 25*, 408-417.

## About the Author

*Yin Ling CHEUNG earned her Ph.D. in Linguistics from Purdue University. She is Assistant Professor and MA (Applied Linguistics) Programme Leader at the English Language and Literature Academic Group, National Institute of Education, Nanyang Technological University, Singapore. She teaches graduate and undergraduate courses in second language writing.*