The International Research Foundation
for English Language Education

Title of Project:
Exploring Automated Corrective Feedback in EFL
Writing Classroom Context: A Case Study of Criterion

Researcher:
Giang Hoang
University of Melbourne
lgiang8380@gmail.com

Dissertation Supervisors:
Associate Prof. Ute Knoch
University of Melbourne

Giang Hoang

Final Report

Motivation for the Research
Automated writing evaluation (AWE) systems are increasingly used in classroom settings to provide formative feedback to learners. This study emerged from a concern that the increasingly prominent presence of automated writing instructional programs in numerous teaching and learning contexts is not accompanied by adequate understanding of their impact on students' learning to write in a second language. While corporate funded researchers are in a better position to access big data to study how (well) a system scores student essays, there is still a scarcity of research evidence about the impact of automated feedback on accuracy development or writing/revision practices and a lack of longitudinal studies into learners' engagement with automated feedback. Therefore, the current study was conducted to seek insights into learners' actual interaction with automated form-focused feedback through learners' perspectives, which produced implications for both service providers and end-users of automated writing instructional programs.

Research Questions
1. What is the nature of *Criterion* automated corrective feedback (ACF)?
2. How accurate is *Criterion*'s automated corrective feedback?
3. How do EFL students engage with and respond to *Criterion* ACF?
4. How does EFL students' accuracy in L2 writing change after a semester's access to and use of *Criterion* ACF?
5. What are students' perceptions of *Criterion* ACF?

Research Methodology
The study adopted a pre-post quasi-experimental design to examine the effectiveness of ACF as a formative assessment tool for L2 learners' writing development. Participants included 81 English majors in a teacher-training degree in the English Department of a university in Hue, central Vietnam. The students, sub-divided into the experimental and control groups, were undertaking a 15-week writing course during which both groups had three practice sessions to compose essays to three writing prompts. The control group wrote their essays on paper and submitted them to the instructor for feedback. The experimental group, however, had

access to the automated writing instructional program called *Criterion* during their practice sessions, so they were able to revise their drafts in response to its feedback before submitting revised drafts to the teacher for feedback.

Beside test essays, other data included the following: (1) first and revised drafts from *Criterion* practice sessions; (2) recorded think-aloud protocols (TAP), which were conducted with 14 students as they revised essays using *Criterion* corrective feedback; (3) follow-up interviews with four students, and (4) end-of-term focus group interviews. Error analysis was conducted on the test scripts of the control and experimental groups to investigate any changes in accuracy over time. Two accuracy measures were used: (1) a holistic measure based on Foster and Wigglesworth's (2016) weighted clause ratio approach, and (2) analytic measures of accuracy in using specific grammatical morphemes following an obligatory occasion analysis. Findings from the quantitative pre- and posttest analyses were triangulated with the qualitative data, showing how the students engaged with the feedback from *Criterion* and implemented it in their revision practices. Students' revisions were analyzed for the type of changes made in response to *Criterion* corrective feedback. Analyses of TAP recordings employed Schmidt's (1993) key concept of "noticing" to shed light on students' cognitive engagement with the automated feedback and to clarify revision practices found. Further triangulation came from students' perceptions of automated feedback in TAP follow-up and focus group interviews that were thematically analyzed.

Summary of Findings

Criterion was found to perform differently across different error types. Overall, the automated corrective feedback was satisfactorily accurate with 21 error types exceeding the 80% precision threshold. Findings on students' engagement with and subsequent revisions using *Criterion* ACF revealed more perfunctory than substantive processing of the feedback. It was also found that unhelpful feedback tended to initiate more extensive engagement among the learners. Following their engagement, students performed mostly local textual operations to revise their texts and were modestly successful at correcting the errors *Criterion* pointed out to them, with 56.2% success rate and 7.1% retention of their original texts in response to false positives. Generally, the surface-level nature of the feedback and students' superficial engagement with it, along with simple and straightforward textual operations suggest a strong connection between the nature of the feedback, the types of errors addressed, and the way in which students responded to the feedback.

Over the time frame for the study, there was no significant treatment effect of the use of *Criterion* ACF on experimental students' accuracy across the five morphemes for which they received most error tags from *Criterion*, including article, third person singular, copula, plural, and comma usage. Their overall accuracy did not improve significantly either.

The qualitative analyses of the focus group interviews and stimulated recalls revealed students' overall positive perceptions about *Criterion* ACF. Their satisfaction with the feedback was thanks to *Criterion*'s immediacy, systematicity, and comprehensiveness in generating feedback. Students also discussed issues that caused concerns about *Criterion* ACF, including the lack of elaboration in metalinguistic explanations and absence of feedback on more complex grammatical issues. The findings also pointed out learners' varied expectations about how explicit the feedback should be and their evaluations of teacher compared to automated corrective feedback.

Overall, the validation of *Criterion* ACF as a learning and assessment tool in the EFL writing classroom reveals a mixture of support for its use and rebuttal evidence against its use. *Criterion* was able to address EFL learners' needs for surface-level errors, but it still lacked coverage of some major issues in the students' L2 writing. It can be praised for facilitating revising, as well as self-regulatory writing strategies, and triggering noticing among the students; however, *Criterion*'s approach to feedback generation was not pedagogically based, resulting in a lack of meaningful engagement with the feedback. Overall, despite students' positive feelings about *Criterion* ACF, reservations about its value remain due to learners' middling revision success rates and the absence of significant intervention or retention effects of the use of *Criterion* ACF on their accuracy gains over the period of study.

Implications

The findings extend our understanding about students' engagement and use of the automated corrective feedback by adding a missing piece in the research on EFL learners' use of *Criterion* ACF. Qualitative data from think-aloud protocols and stimulated recalls unveiled the underlying mechanisms that moderate learners' final revising decisions and somewhat invalidate certain findings in earlier research which assumed that the absence of any change to a flagged error indicates the learners' non-use of the feedback.

The study's main pedagogical implications relate to formative feedback practices in the classroom, including the need to supplement *Criterion* automated feedback with teacher feedback to support L2 writing instruction and classroom-based assessment. Specifically, writing instructors can provide oral feedback sessions during meeting hours so that learners can bring up clarification questions after they have engaged with *Criterion* ACF. Furthermore, strategy training sessions can be embedded in classes during which learners learn cognitive/metacognitive strategies for revisions or share reference sources they find most helpful to seek further information for the error codes received. The instructor should also be open to the current shortcomings of *Criterion* ACF.

From developers' perspectives, *Criterion* corrective feedback should be designed to be more adaptable to focus learners' attention on the relevant issues for their developmental stage. Solutions include making *Criterion* feedback more flexible in terms of toggling between its error types or feedback options to maximize the benefits for a larger number of learners. In addition, more lenient and flexible treatment of writing mechanics issues should be considered to accurately reflect the dynamic nature and changing realities of writing.

References

Activity 4: Complexity in oral vs. written language. (n.d.). Retrieved from https://carla.umn.edu/learnerlanguage/spn/comp/activity4.html?fbclid=IwAR1LSZQ AsiNWCVo2YgO8hSbMcDH1vw-hrPzwaasxXcnrj-W4kHgtHvX7MZk

Attali, Y. (2004). **Exploring the feedback and revision features of Criterion.** *Journal of Second Language Writing*, *14*, 191-205.

Attali, Y., & Burstein, J. (2006). Automated essay scoring with e-rater® V.2. *Journal of Technology, Learning, and Assessment, 4*(3), 1-31.

Barzanji, A. (2016). *Identifying the most common errors in Saudi university students' writing: Does the prompt matter?* (Unpublished master's thesis). University of Central Florida, Orlando, FL.

Ben-Simon, B., & Bennett, R. E. (2007). Toward more substantively meaningful automated essay scoring. *The Journal of Technology, Learning, and Assessment, 6*(1), 1-47.

Benson, S., & DeKeyser, R. (2018). Effects of written corrective feedback and language aptitude on verb tense accuracy. *Language Teaching Research, 23*(6), 702-726.

Bitchener, J. (2012). A reflection on 'the language learning potential' of written CF. *Journal of Second Language Writing, 21*(4)*,* 348-363.

Bitchener, J. (2017). Why some learners fail to benefit from written corrective feedback. In H. Nassaji & E. Kartchava (Eds.), *Corrective feedback in second language teaching and learning: Research, theory, applications, implications* (pp. 129-140)*.* New York, NY: Taylor & Francis.

Bitchener, J., & Ferris, D. R. (2012). *Written corrective feedback in second language acquisition and writing*. New York, NY: Routledge.

Bitchener, J., & Knock, U. (2009). The relative effectiveness of different types of direct written corrective feedback. *System, 37*(2), 322-329.

Bitchener, J., & Storch, N. (2016). *Written corrective feedback for L2 development*. Bristol, UK: Multilingual Matters.

Bitchener, J., Young, S., & Cameron, D. (2005). The effects of different types of corrective feedback on ESL student writing. *Journal of Second Language Writing, 14*(3), 191-205.

Burstein, J., & Chodorow, M. (1999, June). Automated essay scoring for nonnative English speakers. In M. B. Olsen (Chair), *Computer mediated language assessment and evaluation in natural language processing.* Symposium conducted at the meeting of Association for Computational Linguistics, Stroudsburg, PA. Retrieved from http://portal.acm.org/citation.cfm?id=1598834&picked=prox&cfid=10764046&cftoke n=66767680

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

Burstein, J., Chodorow, M., & Leacock, C. (2003, August). *Criterion^{SM} online essay evaluation: An application for automated evaluation of student essays.* In J. Riedl, & R. Hill (Eds.), *Proceedings of the Fifteenth Annual Conference on Innovative Applications of Artificial Intelligence* (pp. 3-10). Acapulco, Mexico: AAAI Press.

Chandler, J. (2003). The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of L2 student writing. *Journal of Second Language Writing, 12*(3)*,* 267-296.

Chapelle, C. A., Cotos, E., & Lee, J. Y. (2015). Validity arguments for diagnostic assessment using automated writing evaluation. *Language Testing*, *32*(3), 385-405. https://doi.org/10.1177/0265532214565386

Chen, C. E., & Cheng, W. (2006). *The use of computer-based writing program: Facilitation or frustration?* Retrieved from ERIC database. (ED492078)

Chen, H. H., Chiu, S. T., & Liao, P. (2009). Analyzing the grammar feedback of two automated writing evaluation systems: MY Access and Criterion. *English Teaching and Learning, 33*(2), 1-43.

Chodorow, M., & Burstein, J. (2004). *Beyond essay length: Evaluating e-rater®'s performance on TOEFL essays* (Report No. 73). Princeton, NJ: ETS.

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*(3), 419-436. https://doi.org/10.1177/0265532210364391

Choi, J. (2010). *The impact of automated essay scoring (AES) for improving English language learners' essay writing* (Unpublished doctoral dissertation). University of Virginia, Charlottesville, Virginia.

Cohen, A. D. (2000). Exploring strategies in test-taking: Fine-tuning verbal reports from respondents. In G. Ekbatani & H. Pierson (Eds.), *Learner-directed assessment in ESL* (pp. 127-150). Mahwah, NJ: Lawrence Erlbaum.

Coniam, D. (2009). Experimenting with a computer essay-scoring program based on ESL student writing scripts. *ReCALL, 21*(2), 259-279. https://doi.org/10.1017/S0958344009000147

Cotos, E. (2011). Potential of automated writing evaluation feedback. *CALICO Journal, 28*(2), 420-459.

Creswell, J. W. (2013). *Qualitative inquiry & research design: Choosing among five approaches*. (3rd ed.). Thousand Oaks, CA: SAGE Publications.

Cumming, A., Kantor, R., Baba, K., Eouanzoui, K., Erdosy, U., & James, M. (2006). Analysis of discourse features and verification of scoring levels for independent and

integrated prototype written tasks for the new TOEFL [Monograph]. *ETS TOEFL Monograph Series, MS-30*. Princeton, NJ: ETS.

Dam, P. (2001, February). *Old habits die hard: Persistent errors in English written by Vietnamese speakers*. Paper presented at the National Association for Bilingual Education, Phoenix, AZ.

Dao, D. V. (2018). Some Vietnamese students' problems with English grammar: A preliminary study. *Hawaii Pacific University TESOL Working Paper Series, 6*(2), 37-55. Retrieved from https://www.hpu.edu/research-publications/tesol-working-papers/2008-fall/6.2-05-Dao.pdf

Deane, P. (2013). On the relation between automated essay scoring and modern views of the writing construct. *Assessing Writing, 18*(1), 7-24.

Dikli, S. (2006). *Automated essay scoring in an English as a second language setting* (Doctoral dissertation, Florida State University). Retrieved from http://etd.lib.fsu.edu/theses_1/available/etd-07052007-152924/unrestricted/sd_dissertation.pdf

Dikli, S. (2010). The nature of automated essay scoring feedback. *CALICO Journal, 28*(1), 99-134.

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.

Dornyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Routledge. ProQuest Ebook Central, http://ebookcentral.proquest.com/lib/unimelb/detail.action?docID=257290.

Dulay, H., & Burt, M. (1974). Natural sequences in child second language acquisition. *Language Learning 24,* 37-53.

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

Ellis, R. (2009). A typology of written corrective feedback types. *ELT Journal*, *63*(2), 97-107.

Ellis, R. (2010). Epilogue: A framework for investigating oral and written corrective feedback. *Studies in Second Language Acquisition, 32*, 335-349. https://doi.org/10.1017/S0272263109990544

Ellis, R., & Barkhuizen, G. (2005). *Analysing learner language*. Oxford, UK: Oxford University Press.

Ellis, R., Sheen, Y., Murakami, M., & Takashima, H. (2008). The effects of focused and unfocused written corrective feedback in an English as a foreign language context. *System, 36*, 353-371.

Ericsson, K., & Simon, H. (1993). *Protocol analysis: Verbal reports as data* (Rev. ed.). Cambridge, MA: MIT Press.

Evans, N. W., Hartshorn, K. J., Cox, T. L., & de Jel, T. M. (2014). Measuring written linguistic accuracy with weighted clause ratios: A question of validity. *Journal of Second Language Writing, 24,* 33-50. https://doi.org/10.1016/j.jslw.2014.02.005

Fang, Y. (2010). Perceptions of the computer-assisted writing program among EFL college learners. *Educational Technology & Society, 13*(3), 246-256.

Farrokhi, F., Sattarpour, S. (2012). The effects of direct written corrective feedback on improvements of grammatical accuracy of high-proficient L2 learners. *World Journal of Education, 2*(2), 49-57.

Feng, H. H., Saricaoglu, A., & Chukharev-Hudilainen, E. (2016). Automated error detection for developing grammar proficiency of EFL learners. *CALICO Journal, 33*(1), 49-70.

Ferris, D. R. (1999). The case for grammar correction in L2 writing classes. A response to Truscott (1996). *Journal of Second Language Writing*, 8, 1–11.

Ferris, D. R. (2003). *Response to student writing: Implications for second language students*. Mahwah, NJ: Lawrence Erlbaum Associates.

Ferris, D. R., Liu, H., Sinha, A., & Senna, M. (2013). Written corrective feedback for individual L2 writers. *Journal of Second Language Writing*, 22, 307-329.

Ferris, D., & Roberts, B. (2001). Error feedback in L2 writing classes: How explicit does it need to be? *Journal of Second Language Writing, 10*(3), 161-184.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London, UK: SAGE Publications Ltd.

Foltz, P. W., Laham, D., & Landauer, T. K. (1999). The Intelligent Essay Assessor: Applications to educational technology. *Interactive Multimedia Educational Journal of Computer-Enhanced Learning, 1*(2). Retrieved from http://imej.wfu.edu/articles/1999/2/04/

Foster, P., & Wigglesworth, G. (2016). Capturing accuracy in second language performance: The case for a weighted clause ratio. *Annual Review of Applied Linguistics, 36*, 98-116.

Fuchs, R., & Werner, V. (2018). Introduction. *International Journal of Learner Corpus Research. 4*(2), 143-163.

Grindsted, A. (2004). Interactive resources used in research interviewing. *Hermes, Journal of Linguistics, 33*, 117-144.

Han, N., Chodorow, M., & Leacock, C. (2006). Detecting errors in English article usage by non-native speakers. *Natural Language Engineering, 12*(2), 115-129. https://doi.org/10.1017/S1351324906004190

Han, Y., & Hyland, F. (2015). Exploring learner engagement with written corrective feedback in a Chinese tertiary EFL classroom. *Journal of Second Language Writing*, *30*, 31-44.

Heift, T., & Hegelheimer, V. (2017). Computer-assisted corrective feedback and language learning. In H. Nassaji and E. Kartchava (Eds.), *Corrective feedback in second language teaching and learning: Research, theory, applications, implications* (pp. 51-65). New York, NY. Higgins, D., Burstein, J., & Attali, Y. (2006). Identifying off-topic student essays without topic-specific training data. *Natural Language Engineering, 12*(2), 145-159. https://doi.org/10.1017/S1351324906004189

Hoang, G. T. L., & Kunnan, A. (2016). Automated essay evaluation for English language learners: A case study of *MY Access. Language Assessment Quarterly, 13*(4), 359-376.

Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. *Language Teaching, 39*, 83-101. https://doi.org/10.1017/S0261444806003399

Hyland, K., & Hyland, F. (2019). Contexts and issues in feedback on L2 writing. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (2nd ed.) (pp. 1-22). Cambridge, UK: Cambridge University Press.

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. 79-92). Logan, UT: Utah State University Press.

Kang, E., & Han, Z. (2015). The efficacy of written corrective feedback in improving L2 written accuracy: A meta-analysis. *The Modern Language Journal, 99*(1), 1-18.

Kellogg, R. T., Whiteford, A. P., & Quinlan, T. (2010). Does automated feedback help students learn to write? *Journal of Educational Computing Research, 42*(2), 173-196.

Kelly, A. (2005). General models for automated essay scoring: Exploring an alternative to the status quo. *Journal of Educational Computing Research, 33*(1), 101-113.

Kim, H. R., & Bowles, M. (2019). How deeply do second language learners process written corrective feedback? Insights gained from think-alouds. *TESOL Quarterly, 53*(4), 913-938. https://doi.org/10.1002/tesq.522

Koizumi, R., In'nami, Y., Asano, K., & Agawa, T. (2016). Validity evidence of *Criterion* for assessing L2 writing proficiency in a Japanese university context. *Language Testing in Asia, 6*(5), 1-26. https://doi.org/10.1186/s40468-016-0027-7

Kukich, K. (2000, September). Beyond automated essay scoring. In M. A. Hearst (Ed.), *The debate on automated essay grading* (pp. 22-27). *IEEE Intelligent Systems, 15*(5), 22-27.

Lai, Y. (2010). Which do students prefer to evaluate their essays: Peers or computer program. *British Journal of Educational Technology, 41*(3), 432-454. https://doi.org/10.1111/j.1467-8535.2009.00959.x

Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics, 24*(7), 590-619.

Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS.* New York, NY: Routledge.

Lavolette, E., Polio, C., & Kahng, J. (2015). The accuracy of computer-assisted feedback and students' responses to it. *Language Learning & Technology, 19*(2), 50–68.

Leacock, C., Chodorow, M., Gamon, M., & Tetreault, J. (2014). *Automated grammatical error detections for language learners* (2nd. ed.). Synthesis Lectures on Human Language Technologies. San Rafael, CA: Morgan and Claypool.

Lee, H., Li, J., & Hegelheimer, V. (2012, September). *The impact of Criterion on error reduction: A longitudinal study*. Paper presented at TSLL conference, Ames, IA.

Leow, R. P., & Morgan-Short, K. (2004). To think aloud or not to think aloud: The issues of reactivity in SLA research methodology. *Studies in Second Language Acquisition, 26*(1),35-57. https://doi.org/10.1017/S0272263104261022

Li, J., Lee, H. W., Lee, J. Y., Karakaya, K., & Hegelheimer, V. (2011). *The influence of using Criterion on students' error correction in writing*. Paper presented at the Second Language Research Forum, Ames, IA.

Li, J., Link, S., & Hegelheimer, V. (2015). Rethinking the role of automated writing evaluation (AWE) feedback in ESL writing instruction. *Journal of Second Language Writing, 27*, 1-18.

Liu, S., & Kunnan, A. J. (2016). Investigating the application of automated writing evaluation to Chinese undergraduate English majors: A case study of *WriteToLearn*. *CALICO Journal, 33*(1), 71-91

Lutovich, D., & Chan, J. F. (2000). *Just commas: 9 basic rules to master comma usage.* San Anselmo, CA: Advanced Communications Designs, Inc.

Manchón, R. M. (2011). Situating the learning-to-write and writing-to-learn dimensions of L2 writing. In R. M. Manchón (Ed.), *Learning-to-write and writing-to-learn in an additional language* (pp. 3-15). Philadelphia, PA: John Benjamins.

McGee, T. (2006). Taking a spin on the Intelligent Essay Assessor. In P. F. Ericsson & R. H. Haswell (Eds.), *Machine scoring of student essays: Truth and consequences* (pp. 79-92). Logan, UT: Utah State University Press.

Mehrabi-Yazdi, O. (2018). Short communication on the missing dialogic aspect of an automated writing evaluation system in written feedback research. *Journal of Second Language Writing, 41*, 92-97

Mikulas, C., & Kern, K. (2006, April). *A comparison of the accuracy of automated essay scoring using prompt-specific and prompt-independent training*. Paper presented at

the Annual Meeting of the American Educational Research Association (AERA), San Francisco, CA.

Min, H. (2006). The effects of trained peer review on EFL students' revision types and writing quality. *Journal of Second Language Writing, 15*(2), 118-141.

Mohsen, M. A., & Alshahrani, A. (2019). The effectiveness of using a hybrid mode of automated writing evaluation system on EFL students' writing. *Teaching English with Technology, 19*(1), 118-131.

Nichols, P. D. (2004, April). *Evidence for the interpretation and use of scores from an automated essay scorer*. Paper presented at the Annual Meeting of the American Educational Research Association (AERA), San Diego, CA.

Norris, J., & Ortega, L. (2003). Defining and measuring SLA. In C. J. Doughty & M. H. Long (Eds.). *The handbook of second language acquisition* (pp. 717-761). Malden, MA: Blackwell Publishing.

Page, E. B. (1968). The use of the computer in analyzing student essays. *International Review of Education, 14*(2), 210-225. https://doi.org/10.1007/BF01419938

Page, E. B., & Petersen, N. S. (1995). The computer moves into essay grading: Updating the ancient text. *Phi Delta Kappan, 76*(7), 561-565.

Peterson, S., Childs, R. M., & Kennedy, K. (2004). Written feedback and scoring of sixth-grade girls' and boys' narrative and persuasive writing. *Assessing Writing, 9*(2), 160-180.

Philp, J., & Duchesne, S. (2016). Exploring engagement in tasks in the language classroom. *Annual Review of Applied Linguistics, 36*, 50-72.

Phuket, P. R. N., & Othman, N. B. (2015). Understanding EFL students' errors in writing. *Journal of Education and Practice, 6*(32), 99-106.

Pica, T. (1982). Adult acquisition of English as a second language under different conditions of exposure. *Language Learning, 33*(4), 465-497.

Polio, C. (2012). The relevance of second language acquisition theory to the written error correction debate. *Journal of Second Language Writing, 21*, 375-389.

Polio, C. G. (1997). Measures of linguistic accuracy in second language writing research. *Language Learning, 47*(1), 101-143.

Polio, C., & Friedman, D. A. (2017). *Understanding, evaluating, and conducting second language writing research*. New York, NY: Routledge.

Powers, D. E., Burstein, J. C., Chodorow, M., Fowles, M. E., & Kukich, K. (2000). *Comparing the validity of automated and human essay scoring* (Research Report No. 98-08aR). Princeton, NJ: Educational Testing Service.

Powers, D. E., Burstein, J. C., Chodorow, M., Fowles, M. E., & Kukich, K. (2002). Stumping *e-rater*: Challenging the validity of automated essay scoring. *Computers in Human Behavior, 18*(2), 103-134.

Qi, D. S., & Lapkin, S. (2001). Exploring the role of noticing in a three-stage second language writing task. *Journal of Second Language Writing, 10*, 277-303.

Quinlan, T., Higgins, D., & Wolff, S. (2009). *Evaluating the construct-coverage of the e-rater scoring engine* (Research Report No. RR-09-01). Princeton, NJ: Educational Testing Service.

Ranalli, J. (2018). Automated written corrective feedback: How well can students make use of it? *Computer Assisted Language Learning, 31*(7), 653-674. https://doi.org/10.1080/09588221.2018.1428994

Ranalli, J., Link, S., & Chukharev-Hudilainen, E. (2017). 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, 37*(1), 8-25. https://doi.org/10.1080/01443410.2015.1136407

Robinson, *P. (**1995**). Attention, memory and "noticing" hypothesis. Language Learning, 45(2), 283-331.* https://doi.org/10.1111/j.1467-1770.1995.tb00441.x.

Rudner, L. M., Garcia, V., & Welch, C. (2006). An evaluation of the *IntelliMetric* essay scoring system. *The Journal of Technology, Learning, and Assessment, 4*(4). Retrieved from http://ejournals.bc.edu/ojs/index.php/jtla/article/view/1651/1493

Rutherford, A. (2000). *Introducing ANOVA and ANCOVA: A GLM Approach*. London, UK: SAGE Publications.

Sachs, R., & Polio, C. (2007). Learners' uses of two types of written feedback on a L2 writing revision task. *Studies in Second Language Acquisition, 29,* 67-100. https://doi.org/ 10.1017/S0272263107070039

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*, 129-158.

Schmidt, R. (1993). Awareness and second language acquisition. *Annual Review of Applied Linguistics, 13*, 206-226.

Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 1–63). Honolulu, HI: Second Language Teaching and Curriculum Center, University of Hawai'i.

Schmidt, R. (2001). Attention. In P. Robin (ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge, UK: Cambridge University Press.

Seliger, H. (1984). Processing universals in second language acquisition. In F. R. Eckman, L. H. Bell, & D. Nelson (Eds.), *Universals of second language acquisition* (pp. 36-47). Rowley, MA: Newbury House.

Sengupta, S. (1998). **From text revision to text improvement: A story of secondary school composition.** *RELC Journal, 29*(1), 110-137.

Sheen, Y. (2007). The effect of focused written corrective feedback and language aptitude on ESL learners' acquisition of articles. *TESOL Quarterly, 41*(2), 255-283.

Sheen, Y., Wright, D., & Moldawa, A. (2009). Differential effects of focused and unfocused written correction on the accurate use of grammatical forms by adult ESL learners. *System, 37*(4), 556-569.

Shermis, M., Burstein, J., Elliot, N., Miel, S., & Foltz, P. (2015). Automated writing evaluation: A growing body of knowledge. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds), *Handbook of writing research* (pp. 395–409). New York, NY: Guilford Press.

Shintani, N., & Ellis, R . (2013). The comparative effect of metalinguistic explanation and direct written corrective feedback on learners' explicit and implicit knowledge of the English indefinite article. *Journal of Second Language Writing*, *22*(3), 286-306. https://doi.org/10.1016/j.jslw.2013.03.011 10.1016/j.jslw.2013.03.011

Smagorinsky, P. (1998). Thinking and speech and protocol analysis. *Mind, Culture, and Activity: An International Journal, 5*(3), 157-177.

Stake, R. E. (1995). *The art of case study research*. Thousand Oaks, CA: SAGE Publications.

Stevenson, M. (2016). A critical interpretative synthesis: The integration of automated writing evaluation into classroom writing instruction. *Computers and Composition, 42, 1-16*.

Stevenson, M., & Phakiti, A. (2014). The effects of computer-generated feedback on the quality of writing. *Assessing Writing*, *16*, 51-65.

Storch, N. (2009). The impact of studying in a second language (L2) medium university on the development of L2 writing. *Journal of Second Language Writing, 18*, 103-118.

Storch, N. (2018). Written corrective feedback from sociocultural theoretical perspectives: A research agenda. *Language Teaching, 51*(2), 262-277. https://doi.org/10.1017/S0261444818000034

Storch, N., & Tapper, J. (2009). The impact of an EAP course on postgraduate writing. *Journal of English for Academic Purposes, 8*(3), 207-223.

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. https://doi.org/10.1017/S0272263109990532

Strause J., Kaufman, L., & Stern, T. (2014). *The blue book of grammar and punctuation* (11ed.). San Francisco, CA: Jossey-Bass.

Sumalinog, G. G. (2018). Common grammatical errors of the high school students: The teacher's perspective. *International Journal of Research and Science & Management, 5*(10), 69-74.

Sun, S. H. (2013). *Written corrective feedback: Effects of focused and unfocused grammar correction on the case acquisition in L2 German* (Doctoral dissertation, The University of Kansas). Retrieved from https://kuscholarworks.ku.edu/bitstream/handle/1808/12284/Sun_ku_0099D_12694_DATA_1.pdf?sequence=1

Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics*, *16*(3), 371-391.

Tetreault, J. R., & Chodorow, M. (2008, August). The ups and downs of preposition error detection in ESL writing. In D. Scott, & H. Uszkoreit (Eds.), *Proceedings of the 22nd International Conference on Computational Linguistics: Vol. 1* (pp. 865-872). Manchester, UK: Association for Computational Linguistics.

Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning*, *46*, 327-369.

Truscott, J. (2004). Dialogue: Evidence and conjecture on the effects of correction: A response to Chandler. *Journal of Second Language Writing*, 13(4), 337–343.

Van Beuningen, C. G., De Jong N., & Kuiken, F. (2012). Evidence on the effectiveness of comprehensive error correction in second language writing. *Language Learning*, *62*(1), 1-41. https://doi.org/10.1111/j.1467-9922.2011.00674.x.

Wang, F., & Wang, S. (2012). A comparative study on the influence of automated evaluation system and teacher grading on students' English writing. *Procedia Engineering, 29*, 993-997.

Wang, P. (2015). Effects of an automated writing evaluation program: Student experiences and perceptions. *Electronic Journal of Foreign Language Teaching, 12*(1), 79. Retrieved from http://search.proquest.com.ezp.lib.unimelb.edu.au/docview/1792498958?accountid=12372

Warden, C. A. (2000). EFL business writing behaviors in differing feedback environments. *Language Learning, 50*(4), 573-616.

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

Weigle, S. C. (2010). Validation of automated scores of TOEFL iBT tasks against non-test indicators of writing ability. *Language Testing, 27*(3), 335-353. https://doi.org/10.1177/0265532210364406

Williams, R., & Dreher, H. (2005). Formative assessment visual feedback in computer graded essays. In E. B. Cohen & L. Kozminski (Eds.), *Issues in informing science and information technology* (pp. 23-32). Santa Rose, CA: Informing Science Press.

Woodside, A. G. (2010). *Case study research: Theory, methods, practice*. Bingley, UK: Emerald Group Publishing Limited.

Wresch, W. (1993). The imminence of grading essays by computer - 25 years later. *Computers and Compositions, 10*(2), 45-58.

Yang, N. D. (2004, March). Using MY Access in EFL writing. *The proceedings of 2004 International Conference and Workshop on TEFL & Applied Linguistics*. Compiled by Department of Applied English, Ming Chuan University (pp. 550-564). Taipei, Taiwan: The Crane Publishing Co.

Yang, Y., Buckendahl, C. W., Juszkiewicz, P. J., & Bhola, D. S. (2002). A review of strategies for validating computer-automated scoring. *Applied Measurement in Education, 15*(4), 391-412.

Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). Thousand Oaks, CA: SAGE Publications.

Zhang, Z. (2017). Student engagement with computer-generated feedback: A case study. *ELT Journal, 71*(3), 317-328. https://doi.org/10.1093/elt/ccwo89

Zhang, Z., & Hyland, K. (2018). Student engagement with teacher and automated feedback on L2 writing. *Assessing Writing, 36*, 90-102.

Zheng, Y., & Yu, S. (2018). Student engagement with teacher written corrective feedback in EFL writing: A case study of Chinese lower-proficiency students. *Assessing Writing*, 37, 13-24.

Ziegler, W. W. (2006). Computerized writing assessment: Community college faculty find reasons to say 'not yet'. In P. E. Ericsson & R. H. Haswell (Eds.), *Machine scoring of student essays: Truth and consequences* (pp. 79-92). Logan, UT: Utah State University Press.