

THE EFFECTIVENESS OF USING A HYBRID MODE OF AUTOMATED WRITING EVALUATION SYSTEM ON EFL STUDENTS' WRITING

by **Mohammed Ali Mohsen** and **Abdulaziz Alshahrani**

Najran University, Saudi Arabia

mmohsen1976 @ gmail.com

Abstract

Automated Writing Evaluation programs have been used extensively to assist both L2 instructors and learners to get corrective feedback and to score students' final product of writing. Research has found that the AWE programs help in optimizing the writing output. However, little is known about the hybrid mode; use of AWE involving the evaluation of both modes instructors and the AWE program. This paper studies the effects of both modes in developing the students' writing outputs using a small case study of 6 EFL learners. The learners were exposed to both modes where in each mode they undertook two sessions using the program. In the first phase the learners wrote an essay via *MY Access* and then saved their input in the program. In the second session, they revised their essays based on the feedback given from the program. In the hybrid mode, the same students in the second session revised their input as per the instructor's feedback and then continued submitting their essays via *MY Access*. Results found that under the hybrid condition students significantly outscored the learners with the AWE program.

Keywords: automated writing evaluation; corrective feedback; writing; hybrid

1. Introduction

The notion of corrective feedback has been increasingly enhanced by the advent of automated writing evaluation (AWE) software such as *e-rater*, *MY Access*, *Holt Online Scoring*, *BETSY* and *Criterion*. The positive efficacy of such programs has been demonstrated by empirical studies (see Stevenson & Phakiti, 2014 for a comprehensive review). Despite their limitations in detecting writing content errors, they have helped in providing immediate feedback of mechanical errors for students' writing, something a human cannot always do (Lavolette, Polio & Kahng, 2015).

The computer-generated feedback provides comments in the form of cast, meta-linguistics, scoring and/or correction. Instructors may find it hard to give instant feedback for students' problematic areas of their pieces of writing, but AWE can partially do that for certain aspects of the language. According to Lavolette et al., (2015), error codes generated by

Criterion were 75% correct. Besides, Hoang and Kunnan (2016) found 73% precision of error scoring provided by *MY Access*. In fact, the issue of corrective written feedback of AWE programs has been debatable for years. Proponents of such programs, who are frequently affiliated with companies that develop such programs, laud their precision and valuable feedback. On the other hand, opponents of these programs base their criticism on the call of Truscott (1999) to abandon such software as they focus on correcting grammatical errors which could lead to surface learning and that could not foster L2 acquisition as the ultimate goal. In fact, whatever accuracy these programs offer, human intervention is essential to control the limits of the program and to advance the high quality of corrective feedback.

Due to mixed designs of the previous studies, lack of validity of such corrective feedback provided by AWE, diversity of programs' features, and shortage of empirical studies, we cannot draw a firm conclusion of the efficacy of these programs. Therefore, more studies are called for to gain a fine-grained picture about the final product of students' writing mediated by AWE programs.

The current study aims to determine the efficacy of AWE (*MY Access*) in developing students' revision of essay writing and to examine how the students' scores improved from the first draft to the second one in two different writing tasks via a computer-assisted writing affordance.

2. Literature review

Recently, a number of writing programs have been developed to assess students' writing as well as provide formative and summative feedback on their writing. Such programs are known as Automated Essay Scoring (AES) (Shermis & Buretein, 2003) or Automated Writing Evaluation (AWE) (Warschauer & Ware, 2008), Examples of AES/AWE include *e-rater*, *MY Access*, *Holt Online Scoring*, *BETSY* and *Criterion*. AES or (AWE) has been described as computer technology that evaluates and scores written prose with the purpose of saving time, reducing cost, and increasing reliability in the assessment of writing (e.g. Chung & O'Neil, 1997; Hamp-Lyons, 2001; Rudner & Liang, 2002).

However, research into the use of automated applications has yielded inconclusive findings. Some studies have reported positive results (Coniam, 2009) while others have reported negative or mixed results (Lai 2010; Lee et al, 2009; Tuzi, 2004). These contradictory results could be attributed to several factors such as individual writing ability, the pedagogy adopted and the specific automated application affecting the results (Lee et al., 2009). For example, less trained writers faced difficulties in using revision tools and also

novice writers could not access these tools (Kozna & Johnston, 1991). Similarly, learners who used *MY Access* were dissatisfied with the grade they received regarding the accuracy and clarity of feedback on content and the rhetorical aspects of their writing (Chen & Cheng, 2008). In contrast, a number of case studies (e.g., Dmytrenko-Ahrabian, 2008; Ellison, 2007; Ussery, 2007) reported student and teacher's satisfaction with the *Criterion* software.

The majority of studies reviewed in the AWE literature have used *Criterion* to provide immediate feedback and scores on students' writings. According to a systematic review study on the use of AWE to improve L2 writing skills which was conducted by Stevenson and Phakiti (2014), around 33% of their selected studies had used *Criterion* to provide immediate feedback to the students' errors while only 15% of studies used *MY Access*. The overuse of such programs in the literature could be explained by the fact that these programs "provide feedback on both global writing skills and language use" (Stevenson & Phakiti, 2014, p. 52). *Criterion* has the potential to give indirect feedback to errors and also provide suggestions to the correct form (Lavolette et al., 2015). Yet, AWE cannot replace instructors and scoring made by such programs cannot be regarded as accurate as human rating and must be treated with "a critical eye" (Warschauer & Ware, 2006, p. 163). Some errors detected by AWE might be misidentified; in other words, some of the errors identified are not really errors and other errors remain unidentified. For the purpose of the current study, we do not aim to validate AWE scoring. On the other hand, our focus is on the corrective feedback provided by *MY Access* in the form of suggestions given to learners and on how such feedback could improve the students' writing when they revise their works in light of these suggestions. Additionally, AWE has been firstly designed to aid native speakers of English who write English prose in their native language (Li & Kunnan, 2016), and little research has targeted English language learners who are not familiar with proper English terms and not exposed to English speaking environments where the English style is unattainable.

In order to examine the effect of *Criterion* on students' writing, particularly by responding to its feedback, a number of studies have been carried out. Attali's (2004) study, for example, reports the results of a large-scale study based on *Criterion* to provide a holistic essay score; feedback on grammar, usage, mechanics, and style. A total number of 9,275 essays were submitted to *Criterion*, which provided feedback to the students who then submitted a revised essay to the program. Data were analyzed from the first and last (of three) essays submitted by US students in the 6th through the 12th grade during the 2002-2003 school year. An overall measure of grammar, usage, mechanics, and style errors were computed by summing the individual error rates, grammar, usage, mechanics, and style errors for each

essay and divided by the essay length to produce an error-rate. Results suggested that overall scores improved and essay length increased for revised submissions compared to the first submission. Similarly, organization and development scores improved and the participants were able to correct at least some types of errors in subsequent versions of their essays.

Lee et al. (2009) developed a system to provide immediate feedback on EFL students' writing as regards content and organization. A comparison was made between essays written by two groups. The experimental group received feedback from the web-based system and the control group typed their essays directly on the computer. It was found that there was no statistically significant difference between the two groups in essay length, or in the final scores given by two human raters.

El Ebyary and Windeatt (2010) examined the potential positive effect of using automated feedback with the help of *Criterion*. The authors sought to examine the trainees' attitudes towards the novel mode of feedback and also investigated both the process of writing and their final product. Quantitative and qualitative data about feedback practice were collected from 31 instructors and 549 Egyptian trainee EFL teachers using pre-treatment questionnaires, interviews and focus groups. A total number of 24 trainees received computer-based feedback using *Criterion* on two drafts of essays submitted on each of the four topics assigned to participants. Data recorded by the software suggested a positive effect on the quality of students' second drafts, subsequent submissions, and post-treatment questionnaires. Similarly, interviews and focus groups showed a positive effect on the students' attitudes towards feedback. In El Ebyary and Windeatt's study, the improvements in students' writing, however, may have been identified partly or mainly due to the novelty (Hawthorne or experimental) effect (McNeill & Chapman, 2005). The authors also argued that issues of writing organization and content were not sufficiently addressed by *Criterion*, and that the errors in language were mainly addressed by the software.

Studies in real classrooms can yield more valid results. However, such research that examines the effect of automated feedback is scarce. Therefore, this study seeks to fill this gap in this area of investigation. Also, to date there is no study that has looked at how a hybrid form of feedback (i.e. automated and teacher feedback) can improve students' writing, and compare this form of feedback with the only one form of feedback (i.e. automated feedback). The current study aims to fill in this gap and contribute to the literature for this under-researched area in written feedback. The study attempts to address the following research questions:

1. What impact can *MY Access!* home edition feedback have on students' writing improvement?
2. What impact can hybrid-mode feedback have on students' writing?
3. Is there any statistically significant difference between the AWE feedback and hybrid mode feedback in improving the students' writing?
4. What are the students' perceptions about the use of AWE feedback on the improvement of their writing?

3. Methodology

3.1. Design

This study opted to use the case study approach to investigate the efficacy of using feedback provided by an automated writing evaluation program for a number of reasons. First, the use of the automated writing evaluation program (*MY Access* writing) has never been used as a pedagogical tool in the educational system in the Saudi EFL context. In fact, the current study context is very likely to be different from other ESL contexts where such an automated evaluation program was used. Therefore, such a different context merits deep investigation. Second, we aimed to investigate and determine what variables could assist us in conducting an experimental study with a larger number of students in the near future when improvements, if any, in the program could be done based on this case study.

3.2. Participants

Twelve EFL Arab students took part in this study. Their proficiency level was intermediate as determined by the placement test administered by the Department of English, Najran University. The proficiency test used was equal to TOEIC. The participants' age ranged from 22 to 24. They had been learning English for at least eight years, including their study at primary and secondary school. All the participants were studying at level 4 (the second semester of the second year of their BA program in English). They were from two different sections of the same level and they were taught by the same teacher (the second researcher). The participants were enrolled in a writing course that aims at teaching how to write an academic essay. All the participants had never been to an English-speaking country, they just learnt English at school and university.

Purposive sampling was used by the teacher/researcher to select the participants. A multiple case study was utilised to find the similarities and differences among the cases and to

increase the reliability of the outcomes. The selection of the participants was based on their academic performance in the teacher's class as well as their academic grade point average (GPA). The researcher selected those participants whose academic GPA was in the range of 3 and 4 out of 5. Based on the academic description of the institution, this range represents good academic performance. This selection was to ensure that all the participants would have the same level of writing proficiency. The participants were briefed about the purpose of the current study. They were assured that participation was voluntary and that the outcome of the study would not have any effect on their grades. A number of participants had attended the first task and then dropped from the study. Only six students completed the two assigned phases of treatments.

3.3. The software program

The software used to gauge students' corrective writing during the assigned sessions was *MY Access*. It is one of the most well-known AWE programs to assist learners in writing skills. "It is a web-based AWE program that uses the Intelli Metric automated essay scoring system. The software, created by Vantage Learning, provides activities for instructors to develop content ideas, organization, language use, help students see other essays that represent different levels of proficiency to understand evaluation criteria, evaluate and grade writing. The program enables students to write their essays and gives them help options such as word bank, feedback, and scoring. Learners can log in the program with their IDs and start recording their input in a file. They can input their essays and save it for later use. Upon automatic scoring provided by the program, the students can polish their inefficiencies and improve the quality of their writing. Figure 1 and Figure 2 are snapshots from *MY Access* program .

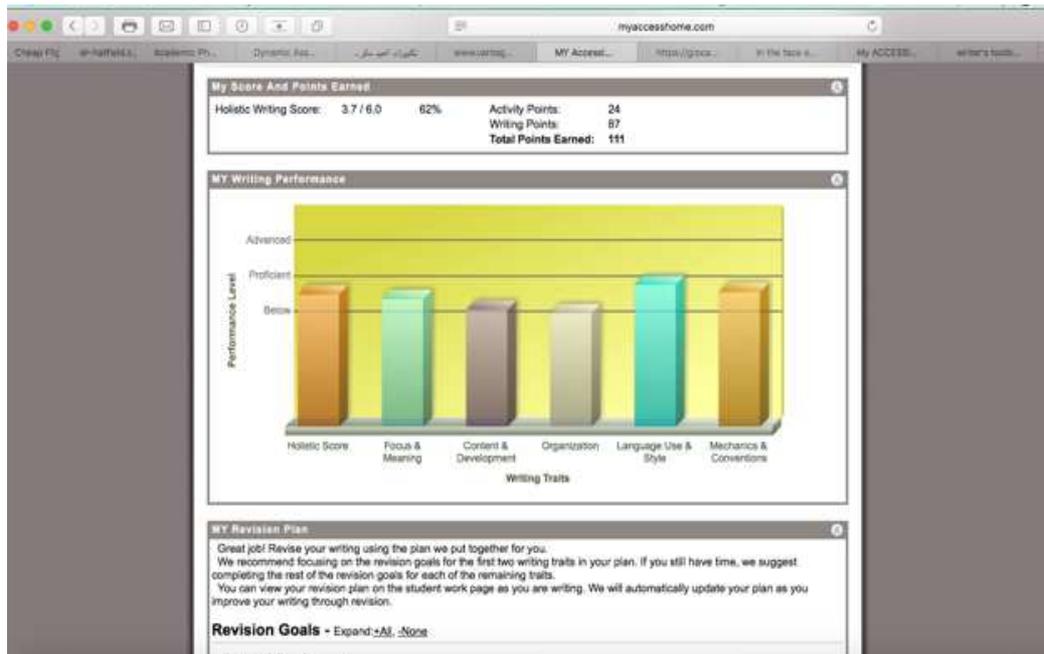


Figure 1. Scoring made by MY ACCESS

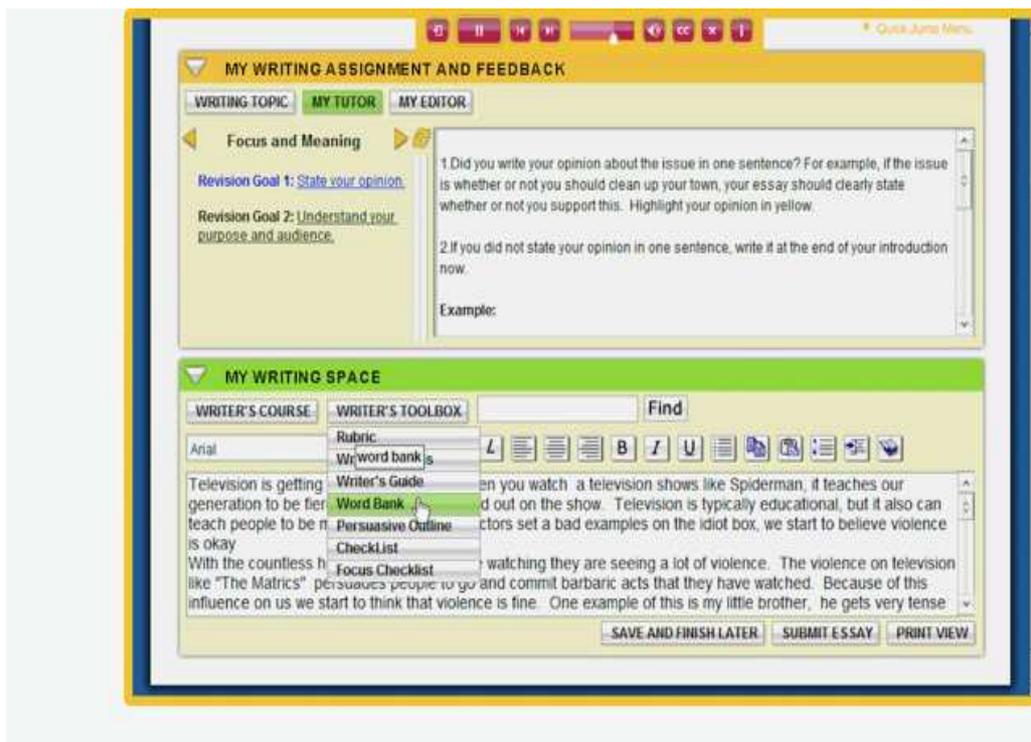


Figure 2. Example of essay written in MY Access with some available functions

3.4. Study Design and Writing Procedures

This study was run over a period of four weeks and included two phases. The second phase (weeks 3 and 4) was based on the findings from phase 1 (weeks 1 and 2). On Day 1 of the study, the six students were brought to the computer lab at the University and were trained to

use *MY Access* writing by the teacher (the second researcher). The teacher explained the different functions of the program and the ways of responding to the feedback. The students were asked to write an essay on a topic chosen from a list. Since the program offers the writer the option to write on a topic from a multiple of proficiency levels (e.g., 8-10, 11-14, 15-18), the students were advised by the researcher to choose a topic from level 4 to suit their proficiency level.

On Day 2 in the first phase, the students used *MY Access* program to write their first draft and then received feedback immediately from the program. The instructor assigned the following topic “The effect of smoking on health”. The students wrote a 3 paragraph-essay on the topic. The instructor chose this topic since the students were familiar with the issue of smoking as it was one of the topics they studied in their textbooks. The students wrote their essays, saved them and then submitted them to the program for feedback. After submission, they immediately received a holistic score out of 6. In the second session, two days after the first session, the students were asked to log into their account and revise their last saved essays. In this session, they were again instructed on how to use *My tutor* to get feedback on their writing content, style, and organization. They were also instructed on how to revise their language errors using *My editor*. Then, the students submitted their second draft and received a holistic score out of 6. The researcher then conducted semi-structured interviews with the students in order to find out how they perceived the program and how it could help them improve their writing.

Based on the findings from phase 1, it was clear that the teacher’s oral feedback intervention was necessary in the areas of the students’ writing content and organization. Thus, a hybrid mode was used in which the students were required to correct their language errors through *My editor* in the program and received feedback from the teacher only on the area of writing content and organization. In the first session of the second phase, the students wrote an essay of their own choice. Each student chose to write a different topic from the range of proficiency levels (8-10, 11-14, 15-18). These topics were of different genres. For example, some were informative (e.g. a good friend), and some were narrative (e.g. your dreams). After writing their essays the students saved and submitted their work and immediately received a holistic score (a maximum score of 6).

In the second session, the students were instructed to revise their saved drafts in the same manner as they did in the first phase. However, in this stage, they were not instructed to use *My tutor* to receive feedback on their writing content, organization and style. It was rather the researcher who provided them with the necessary feedback. Then, the students submitted

their second drafts and received another holistic score. During the two sessions, the instructor took some notes and conducted interviews at the end of the two sessions.

3.5. Interview

To gain insight into the students' perspective in regards to this new automated feedback, semi-structured interviews were used. The semi-structured interview type was chosen for this study because it offers a balance between the flexibility of an open-ended interview and the focus of a structured and restricted interview.

The aim of the interviews was to investigate in greater depth the students' perceptions regarding the new type of feedback by inquiring about their experiences of using it and their preferences over the type of feedback they used to get in their classroom. The questions comprised three different sets. The first set of questions concerned the students' background regarding their learning of writing and receiving feedback. The second set was related to their experience of using their new type of feedback and the difficulties faced. The third set covered the students' preferences regarding this new automated feedback over the one they used to get in their classroom.

The interviews were conducted at the end of the study in a quiet room. This was done in Arabic (the participants' first language) to assist the students to express their ideas and perceptions more easily; thus, allowing for greater investigation by the researchers. The interviews lasted for about twenty minutes with each student and the students' answers were audio-recorded.

4. Results and findings

The results generated by this study were triangulated through administering different data collection tools throughout the study: written tests, semi-structured interviews, observation, note-taking and informal interviews. Quantitative data were collected from the written test while qualitative data were reported from the final semi-structured interviews, the observations, and the informal interviews. Descriptive and inferential statistics were performed to find answers to the research questions of the current study. All the statistical significance level was calculated at .05.

4.1. Quantitative analysis

To answer the first research question which concerns the impact of hybrid feedback on students' writing, students' scores were provided by *MY Access* and are depicted in Table 1 and Table 2 (Note: students are given pseudo-names as to protect their privacy).

Table 1 Students' scores in the First Phase (AWE mode)

Student's name	1 st draft	2 nd draft
Ali	2.7	3.2
Ahmad	1.8	2.3
Hussein	2.0	2.4
Wael	1.9	2.2
Tariq	2.6	3.1
Saad	2.1	2.4

Table 2 Students' scores in the Second Phase (Hybrid Mode)

Student's name	1 st draft	2 nd draft
Ali	2.8	3.5
Ahmad	2.0	2.8
Hussein	2.2	3.1
Wael	2.1	2.9
Tariq	2.7	3.8
Saad	2.3	2.9

Descriptive statistics were used to see the means and standard deviation for both modes. They are summarized in Table 3.

Table 3. Descriptive statistics for students' scores over AWE vs. Hybrid modes

No.	Item	N	M	SD	SE M
1	AWE 1	6	2.18	.37	.15
2	AWE2	6	2.6	.43	.17
3	Hybrid1	6	2.35	.32	.13
4	Hybrid2	6	3.16	.39	.16

Table 1 shows that the students' scores improved from the first session to the second one across the two modes of treatment. For the first phase, means scores increased from the first session ($\bar{x}=2.18$, $SD=.37$) to the second one ($\bar{x}=2.6$, $SD=.43$). A paired-t-test revealed that the improvement from the first session to the second session was significant $t(5)=-10.38$, $p=.000$. Likewise, means scores of the students in the hybrid mode were statistically significant $t(5)=-11.6$, $p=.000$. This result suggests that students' writing would significantly improve when

learners were exposed to the second session of treatment where they can revise their input and make use of the feedback from both the program and the instructor.

In order to determine if there was a statistically significant difference between the two types of modes (AWE and hybrid), a paired sample *t*-test was run. Findings show that the students in the hybrid mode ($\bar{x}=2.75$, $SD=.39$) significantly outscored the same students in the AWE mode ($\bar{x}=2.39$, $SD=.40$, $t(5)= -9.64$, $p=.000$). This reveals that the hybrid mode was beneficial for evaluating students' output and would advance the students' writing skills.

4.2. Qualitative analysis

The data collected from the interviews and observations while the students performing their writing tasks and responding to the feedback provided by the program and the semi-structured interviews provided insight into the students' perceptions and experience of using this new program of providing written feedback. The second researcher interviewed the students about their use of the new program in teaching L2 writing. The findings indicated that it was a new experience for the learners to write an online essay and to get feedback from both the AWE program and the instructor. The learners showed their great interest in *MY Access* program, especially *My editor*. However, in their response to the benefits they got from different functionalities of the program, they mentioned that they did not benefit from the toolbox features such as word bank, although the instructor repeatedly recommended using this feature. This could possibly be explained by the fact that students had little exposure to the new unfamiliar program. Instead, students preferred to use their well-known dictionary apps in their phones to look up new words. Moreover, the participants expressed the difficulty in understanding the feedback on their writing content and organization that is provided by the feature of *My tutor*, except the feedback provided on their writing accuracy that is provided by the feature of *My editor*. In the second phase of the study, in which the teacher intervened and provided feedback on the students' writing content and organization, the students reported that the feedback provided by the teacher (on content and organization) was clearer and dialogic as compared to the feedback provided by the program (*My tutor*) on content and organization.

5. Discussion

The findings reported in this study suggest an obvious improvement in the students' second draft scores during the second phase compared to the second draft scores in the first phase. This can be attributed to the effectiveness of using the hybrid mode on students' final score.

The findings of the current study support the previous findings in that AWE immediate feedback could help students improve the quality of their writing skills to an acceptable level (Attali, 2004; Lavolette et al, 2015) and human intervention could ensure the accuracy of AWE programs. *MY Access* helped the participants polish out the mechanical errors such as spelling, grammar, and punctuation. However, it failed to correct clarity, coherence, and ambiguity of writing which a human can only do. The instructor evaluated the students' output and made sure that the ideas were well-organized, their works were free from ambiguity and the ideas were made crystal clear. This enhances the notion that technology can assist instructors in acquiring second language but we cannot fully rely on it or we cannot replace human instructors (Chapelle, 1999).

The findings from the observation notes and the final interviews could provide an explanation to this claim. The students argued that using *My tutor* could be intimidating as it provided complex instructions. This can obviously be understood given that the feedback on content and organization was both not specific and very long. In fact, this feedback requires the student to go through multiple stages and would need considerable time to complete. This would be difficult for an intermediate level of English proficiency who studies English as a foreign language. The researcher's own observation confirmed the students' perceptions regarding the complexity of instructions provided by *My tutor*. When the participants attempted to use *My tutor*, they could hardly follow the instructions that involved a number of steps. In other words, *My tutor* involves detailed explanations and it refers the students to other activities that may take a long time to complete.

Furthermore, the findings from the interviews and observation notes indicate that feedback provided specifically by *My tutor* is very general and is not tailored to the specific needs of the student's own essay. This is not surprising given the fact that these instructions are provided by a computer, which lacks personal interaction with the learner. This finding corroborates Stevenson and Phakiti's (2014) report about the difficulties of using automated writing systems for providing feedback to meet the learner's specific needs.

In contrast, the feedback provided by the instructor was dialogic and was tailored to each student's own needs. The instructor was able to help overcome the difficulties that the participants faced while completing the writing assignments. In order to further assist the learner, the instructor used the students' mother tongue (Arabic) as needed. The use of Arabic helped overcome difficulties and enabled students to understand different aspects of writing including organization and content.

6. Conclusions, limitations and suggestions for the future research

The study findings prove that the use of the software program can help students improve their writings from the first session to the second one in the two scenarios. The students benefited much from the hybrid mode where the instructor gives his/her feedback more than the program's feedback. This suggests that L2 instructors are advised to delay corrective feedback from the program but to give their own one. Integration of human instructors may diminish the faults and inefficiency of the AWE programs.

The study has some limitations because of the small sample size. Therefore, future studies should use a large number of participants. Future research should track the students' activities when exposed to writing through AWE to find out how their performance is correlated with students' interactions with the immediate feedback provided by AWE programs, and whether many activities could lead to optimal writing output.

Acknowledgement

The article was made possible through the fund No. NU/SHED/15/127 by the deanship of scientific research, Najran University.

References

- Attali, Y. (2004). Exploring the feedback and revision features of *Criterion*. Paper presented at the Paper presented at the National Council on Measurement in Education, San Diego, CA.
- Chapelle, CA. (1999). Research questions for a CALL research agenda: A reply to Rafael Salaberry. *Language Learning & Technology*, 37(1), 108-113.
- Chen, C. F. E., & Cheng, W. Y. E. C. (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.
- Coniam, D. (2009). A comparison of onscreen and paper-based marking in the Hong Kong public examination system. *Educational Research and Evaluation*, 15(3), 243-263.
- Coniam, D. (2009). Experimenting with a computer essay-scoring program based on ESL student writing scripts. *ReCALL*, 21(2), 259-279.
- Dmytrenko-Ahrabian, M. O. (2008). *Criterion* online writing evaluation service case study: Enhancing faculty attention and guidance. *International Journal of English Studies*, 10(2), 121-142.
- 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.
- Hamp-Lyons, L., & Lumley, T. (2001). Assessing language for specific purposes. *Language Testing*, 18(2), 127-132.
- Herl, H. E., O'Neil Jr, H. F., Chung, G. K. W. K., & Schacter, J. (1999). Reliability and validity of a computer-based knowledge mapping system to measure content understanding. *Computers in Human Behavior*, 15(3-4), 315-333.

- Hoang, G., & Kunnan, A. J. (2016). Automated writing instructional tool for English language learners: A case study of *MY Access*. *Language Assessment Quarterly*, 13(4), 359-376.
- Kozna, R. B., & Johnston, J. (1991). The technological revolution comes to the classroom. *Change*, 23(1), 10-22.
- Lai, Y.-H. (2010). Which do students prefer to evaluate their essays: Peers or computer program. *British Journal of Educational Technology*, 41(3), 432-454.
- Lavolette, E., Polio, C., & Kahng, J. (2014). The accuracy of computer-assisted feedback and students' responses to it. *Language Learning & Technology*, 19(2), 50-68. Retrieved 1 January 2019 from <http://ilt.msu.edu/issues/june2015/lavolettepoliokahng.pdf>
- Lee, C., Wong, K., Cheung, W., & Lee, F. (2009). Web-based essay critiquing system and EFL students' writing: A quantitative and qualitative investigation. *Computer Assisted Language Learning*, 22(1), 57-72.
- 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. <https://doi.org/10.1558/cj.v33i1.26380>
- McNeill, P., & Chapman, S. (2005). *Research Methods*. London: Taylor & Francis Ltd.
- O'Neil, H. F., & Schacter, J. (1997). Test specifications for problem-solving assessment. *CSE Technical Report 463*. Los Angeles: Center for the Study of Evaluation.
- Rudner, L., & Liang, T. (2002). Automated essay scoring using Bayes' Theorem. *The Journal of Technology, Learning, and Assessment (J.T.L.A)*, 1(2), 1-22.
- Shermis, M., & Burstein, J. (2003). *Automated Essay Scoring: A Cross Disciplinary Perspective*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Stevenson, M., & Phakiti, A. (2014). The effects of computer-generated feedback on the quality of writing. *Assessing Writing*, 19(1), 51-65.
- Truscott, J. (1999). The case for grammar correction in L2 writing classes: A response to Ferris. *Journal of Second Language Writing*, 8(2), 111-122.
- Tuzi, F. (2004). The impact of e-feedback on the revisions of L2 writers in an academic writing course. *Computers and Composition*, 21(2), 217-235.
- Ussery, R. (2007). *Criterion Online Writing Evaluation Case Study: Improving Writing Skills across the Curriculum*. Retrieved 11 August 2009 from http://www.ets.org/Media/Products/Criterion/pdf/4216_NCarol3.pdf
- Warschauer, M., & Ware, P. (2008). Learning, change, and power. In J. Coiro, M. Knobel, C. Lankshear, & D. J. Leu (Eds.), *Handbook of Research on New Literacies* (pp. 215-239). Mahwah, NJ: Lawrence Erlbaum Associates.