Enhancing the grammatical accuracy of EFL writing by using an AWE-assisted process approach

Hui-Chuan Liao*
National Kaohsiung University of Applied Sciences, 58, Shenzhong Road, Yanchao District, Kaohsiung 82444, Taiwan, ROC

ARTICLE INFO

Article history:
Received 1 July 2014
Received in revised form 2 February 2016
Accepted 15 February 2016
Available online 12 March 2016

Keywords:
Automated writing evaluation (AWE)
EFL writing
Grammatical accuracy
Noticing
Process writing

ABSTRACT

Several automated writing evaluation (AWE) applications have been developed to facilitate writing improvement. However, few studies have examined the use of an AWE-assisted process-writing approach to facilitate EFL grammatical development. This study examined 63 participants’ grammatical performance in revised and subsequent new essays, learner perceptions and strategies, and possible factors mediating learning in an AWE-assisted process-writing program. Student essays and learner responses to a questionnaire regarding their perceptions on and experiences with using Criterion, an AWE tool, to improve the grammatical aspects of their writing were analyzed. In contrast to the improvement in grammatical performance observed in the revisions of each essay, improvement in the writing of new texts was not observed until the third essay. Furthermore, 18 individual interviews were conducted, and four learner types related to the exercise of learner agency were identified: goal getters, accuracy pursuers, reluctant learners, and late bloomers. Agency appeared to mediate AWE-assisted writing, and the repeated act of language gap noticing and metacognitive strategy use mediated by the process-writing approach appeared to facilitate language modification and longer-term shifts in the students’ initial writing ability, although the effects appeared to occur earlier among the goal getters and accuracy pursuers than among the other learner types.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

In English as a foreign language (EFL) writing classrooms, teachers often encounter difficulty in assisting students in learning the organizational structures vital for effective communication (Flower, 1994) and also addressing repetitive grammatical errors (Hinkel, 2003; Milton, 2006). This situation occurs because EFL learners typically develop language and writing skills in English concurrently (Ferris & Hedgcock, 2005). Writing in an EFL context entails the burden of learning to express thoughts while simultaneously learning English; grammatical errors are unavoidable because of the cognitive demand of performing both tasks at the same time (Hyland, 2003). Therefore, effective writing programs for EFL students at basic and intermediate English levels should encompass both local language features (e.g., grammar and mechanics) and global text features (e.g., content, organization, and coherence) (Chen, Chiu, & Liao, 2009; Ferris & Hedgcock, 2005; Hyland, 2003; Milton, 2006).

* Tel.: +886 919063107, +886 7 3814526x8825.
E-mail address: hliao@kuas.edu.tw.

http://dx.doi.org/10.1016/j.system.2016.02.007
0346-251X/© 2016 Elsevier Ltd. All rights reserved.
In response to the challenge for writing instructors of attending simultaneously to the local and global development of EFL learners, researchers have begun exploring the use of automated writing evaluation (AWE) to assess and facilitate EFL student writing. However, whereas numerous studies have focused on the processes and perspectives of instructors and students, few studies have investigated AWE effects on increasing the level of grammatical accuracy in student essays. Although previous research (Attali, 2004; Chodorow, Gamon, & Tetreault, 2010; Long, 2013) has demonstrated the necessity of establishing a mechanism to ensure that students read and respond to AWE feedback to optimize its effects, it appears that no study has examined the effects of a combined use of a multiple-draft approach and the Criterion system, one of the relatively better known AWE applications in Asian EFL settings, on improving local aspects of EFL student writing. In addition, to the best of this author’s knowledge, no research has investigated whether the language skills learners acquire by reading and responding to AWE feedback are transferable to new writing tasks.

To fill this research gap, the present study incorporated revision as a compulsory component in a writing program in order to empirically investigate the effects of using Criterion feedback in a process-writing approach on enhancing the grammatical accuracy in terms of both revised and new texts. In addition, learner agency, perceptions, and strategies were explored to further determine how learning is mediated in an AWE context. From a pedagogical perspective, this study not only contributes to knowledge regarding the efficacy of incorporating AWE feedback in EFL classrooms but also extends the understanding of possible mediators of successful and unsuccessful learning in a process-writing pedagogy involving AWE.

2. Literature review

This section presents a review of (1) the provision of corrective feedback for second language (L2) learners, (2) the merits and limitations of AWE systems, and (3) the effects of using AWE on reducing grammatical errors.

2.1. Corrective feedback

Hyland and Hyland (2006) indicated that formal accuracy in writing should be achieved to prepare EFL learners for academic success and careers. To enhance grammatical accuracy, EFL writing teachers often use corrective feedback, which is generally expected and welcomed by learners from cultures in which authoritarian teaching has typically been implemented (Hyland & Hyland, 2006). Dikli and Bleyle (2014) cautioned against not offering corrective feedback because EFL students may be uncomfortable when not receiving feedback on form or may conclude erroneously that their written output is grammatically accurate.

Previous studies have indicated that corrective feedback can benefit learners in revised texts (Ferris, 2006; Truscott & Hsu, 2008) and new texts (Bitchener & Knoch, 2010; Sheen, Wright, & Moldawa, 2009). However, providing clear and specific corrective feedback to a class of EFL students creates a heavy workload (Dikli, 2010), particularly when teachers must also attend to learner development regarding essay organization and content. Finding effective and efficient methods for responding to learner essays without turning instructors into “proofreading slaves” (Milton, 2006, p. 125) is imperative. Therefore, using AWE systems as writing assessment and facilitation tools has become a critical topic explored by EFL researchers and practitioners in recent years. In addition, although learner attitudes toward correction and learner characteristics have been found to determine the success of corrective feedback in non-computer-assisted contexts (Faqeih, 2015; Havranek & Cesnik, 2001), few studies have investigated how these factors might mediate language accuracy in computer-assisted language learning (CALL) environments (Heilt & Schulze, 2007), including AWE-assisted writing classrooms.

2.2. Potential and limitations of AWE

AWE uses artificial intelligence to generate scores for essays and generally provides one or both types of writing feedback: global features and local features. Several higher education institutions in Asia have recently employed AWE systems to enhance learning efficacy and reduce grading and instructional workloads (Chen & Cheng, 2008; Chen et al., 2009; Long, 2013; Otoshi, 2005). Recent studies have indicated certain limitations of AWE systems. First, AWE systems require a large corpus in training; they typically score effectively only when the writing prompts are assigned from their own library (Dikli, 2010). Second, AWE systems tend to exhibit reliability in detecting local errors but are ineffective in identifying global concerns (Attali, Lewis, & Steier, 2012; Skoufaki, 2009). Third, the current technology is limited in its ability to comprehensively detect local errors in L2 texts (Dikli & Bleyle, 2014; Milton, 2006). Fourth, the language used in AWE feedback can be complex and thus overwhelm L2 learners who have not yet developed a sufficient mastery of English and who still require additional modeling and guidance in developing English writing skills (Dikli, 2010; Wang, Shang, & Briody, 2013). Fifth, AWE technology lacks human interaction. From this, issues arise including a lack of individualized and communicative feedback (Chen & Cheng, 2008; Dikli, 2010; Vojak, Kline, Cope, McCarthy, & Kalantzis, 2011; Wang et al., 2013) and potentially confusing hypertext navigation for EFL learners (Shang, 2015; Wang et al., 2013).

Whereas the lack of human interaction is considered a drawback from the perspective of the social aspects of writing, other researchers have considered it a potential benefit for L2 writers. Based on the cognitive information processing dimensions of writing, the sense of objectivity likely relieves some of the anxiety of receiving comments and revising texts in response to AWE feedback (Wang & Goodman, 2012). Moreover, because AWE acts as a tireless tutor in providing instant feedback and grammatical explanation, synchronous communication occurs between the system and users within seconds of
submitting a draft, thereby potentially increasing opportunities for students to practice writing and self-editing. Even the most efficient human reader cannot outperform an AWE system in this regard (Dikli & Bleyle, 2014; Warschauer & Grimes, 2008). This instant feedback allows novice writers to focus on specific linguistic features and subsequently improve local aspects of their writing and build writing confidence (Chen & Cheng, 2008; Milton, 2006).

In addition, the portfolio function of an AWE system enables users to save their drafts online for subsequent retrieval, review, reflection, and revision for further feedback. This cyclical process can transform AWE into a formative assessment tool (Dikli, 2010; Yang, 2010) and extend opportunities for students to utilize metacognitive strategies in the writing-revising process, potentially enabling them to become increasingly autonomous learners (Milton, 2006; Wang & Goodman, 2012). Developing learner autonomy not only enhances learning but also benefits teaching. It helps relieve writing instructors of the burdensome task of repetitively responding to local errors; teachers can capitalize on the saved time by scaffolding writer development in global domains (Lai, 2010; Milton, 2006; Warschauer & Grimes, 2008). However, learner autonomy is not achieved simply through teaching and learning; it is mediated by various interrelated factors, including agency and metacognition. According to Gao and Zhang (2011), autonomy originates from agency and requires metacognitive operations in directing and regulating learning. To elucidate the process of developing autonomy, learners’ use of metacognition and exercise of agency should be explored (Dickinson, 1995; Gao & Lamb, 2011; Wenden, 2002).

In light of both the potential and limitations of AWE, researchers, teachers, and software developers have typically agreed that AWE systems are currently inadequate in meeting the needs of L2 writers as a stand-alone tool (Attali et al., 2012; Educational Testing Service [ETS], 2013; Lai, 2010; Wang et al., 2013). Numerous studies have empirically examined the effectiveness of AWE in supplementing teacher instruction. However, most of these have focused on teacher and learner perceptions and the processes of using various systems, including My Access (e.g., Chen & Cheng, 2008; Grimes & Warschauer, 2010; Lai, 2010), Criterion (El Ebyary & Windeatt, 2010; Spencer & Louw, 2008), and researcher-designed systems (Yang, 2010). The effects of using AWE on enhancing grammatical accuracy have rarely been examined.

### 2.3. Effects of using AWE on enhancing linguistic accuracy

To realize the benefits of using AWE, users must have an AWE tool that can detect local errors and provide feedback effectively. Although numerous studies have assessed AWE systems by comparing machine and human ratings, they have focused on the validity and reliability of large-scale writing assessment and primarily involved native English speakers (e.g., Attali et al., 2012; Stevenson & Phakiti, 2014). Few studies have empirically assessed the ability of AWE systems to detect local errors of L2 writers. Researchers have, furthermore, targeted two AWE tools that are better known and commonly used in Asian EFL classrooms (Chen et al., 2009): Vantage My Access and ETS Criterion. Dikli (2010) and Chen et al. (2009) have compared local errors identified by My Access and human raters. Chen et al. analyzed 119 essay pieces written by university students in Taiwan, reporting that My Access provided many false alarms and had a low accuracy rate of 15%. Dikli analyzed 180 essay pieces written by 12 EFL adult learners attending a 7-week intensive English program in the United States, concluding that My Access 6.0 often provides inappropriate error messages pertaining to local features and does not meet the needs of EFL students, particularly those at basic English levels.

Similarly, Otoshi (2005) assessed the accuracy of Criterion in detecting errors in the categories of verbs, word choice, nouns, articles, and sentence structures by analyzing feedback to essays composed by 28 Japanese adult EFL learners and found its performance unsatisfactory. Chen et al. (2009) examined the tool’s accuracy in detecting errors in the categories of articles, spelling, fragments, run-on sentences, subject-verb agreement, ill-formed verbs, compound words, confused words, and proofread this by analyzing the error feedback on 150 essays written by Taiwanese university students. The accuracy rate of the grammatical component reached 79%. Chen et al. argued that although some of the error feedback messages were inappropriate, most of those regarding local language features are instrumental in enhancing the writing accuracy of Taiwanese EFL learners. They stated that the updated version has strong potential to relieve some of the workload of writing instructors and offers learners additional writing opportunities. Dikli and Bleyle (2014) analyzed 37 essay drafts composed by 14 advanced English as a second language (ESL) university students from an English for academic purposes course. The participants were either first- or second-generation immigrants or “generation 1.5”, with an average of 10.36 years of living in the United States. The analyzed error types included wrong or missing words, ill-formed verbs, proofread this, subject-verb agreement, pronoun errors, garbled sentences, fragments, possessive errors, and run-on sentences. An accuracy rate of 63% was observed. The differences between Chen et al. (2009) and Dikli and Bleyle (2014), in addition to the overall accuracy rates of the partly different categories, included the various linguistic backgrounds of the participants (EFL vs. ESL) and human raters (non-native vs. native). In addition, Han, Chodorow, and Leacock (2006) and Tetreault and Chodorow (2008) have examined the performance of Criterion in detecting article and preposition errors, respectively, in L2 writing, concluding that the precision rates were approximately 90% and 80%, respectively. These system-centric evaluation studies have shown that, on average, if a student accepts all the suggested linguistic feedback provided by Criterion when revising an essay, the number of grammatical errors will decrease in the revised draft.

Long (2013) investigated the effects of using Criterion on reducing the number of surface-level errors in EFL essays. However, no revisions were required after students received AWE feedback. This is most likely because of the lack of a mechanism in the pedagogical design to condition the students to read and respond to the AWE feedback, and so no linguistic improvement was detected. This finding parallels that of Warschauer and Grimes (2008), in which revisions were optional; consequently, two-thirds of the L1 participants did not revise their drafts after receiving Criterion feedback. A similar
phenomenon was also observed in Attali (2004): Without a mandatory revision policy, 23,567 out of 33,171 (i.e., over two-thirds) L1 student essay submissions to the Criterion system were not followed up with revisions.

For one-third of the essays involving revisions in response to Criterion feedback, Attali (2004) compared the first drafts with the resubmissions, observing enhanced L1 writer performance indicated by reduced grammatical error rates, increased essay length, and higher holistic scores in subsequent revisions. Chodorow et al. (2010) compared the number of article errors in the first drafts and mandatory revisions addressing the Criterion feedback, and observed substantial improvement of the L2 participants studying at a university in the United States. In Wang et al. (2013), Chinese EFL learners in Taiwan were required to revise their drafts based on the feedback from another AWE tool, Vantage CorrectEnglish. The researchers observed that using AWE facilitated improvement in EFL writing in both accuracy and autonomy. The findings of these studies illustrate the need for incorporating revision as a necessary process for maximizing the effects of AWE feedback.

Long (2013), Attali (2004), and Chodorow et al. (2010) were the first to explore whether using Criterion feedback enhances the linguistic aspect of writing outcomes. Among them, only Long (2013) examined the phenomenon in an EFL learning context; only Chodorow et al. (2010) required learners to revise their texts; and no study examined whether the effects were also exhibited in subsequent new texts. To expand on the previous research, the present study added revision in the pedagogical design as a mandatory step in investigating the efficacy of using Criterion feedback in a process approach on improving local aspects of EFL essays. This study not only compared student performance between the original and revised drafts of each essay but also among the original drafts of various essays. Furthermore, because previous research has indicated that learner profiles and attitudes influence the success of corrective feedback in non-CALL environments, how these factors might mediate learning in a CALL context was investigated in the present study. Specifically, the following research questions (RQs) were explored:

1. At what point in the AWE-assisted process-writing program does learners’ grammatical performance change, if at all?
2. How does grammatical performance as evaluated by Criterion relate to learner perceptions of the effectiveness of the system and learners’ self-reported metacognitive strategy use?
3. Based on learner narratives, what additional factors mediate learner grammatical development in this AWE-assisted writing program?

3. Methodology

This study employed a 9-week time-series research design. Purposeful and subsequent random sampling was used to recruit participants.

3.1. Participants

The participants comprised 63 students, 15 males and 48 females, from three intact sophomore English writing classes at three different universities in Taiwan, where Mandarin Chinese is the official language. They majored in English, had taken two paragraph writing classes in their freshman year, had no experience in using AWE, and were of upper-elementary to upper-intermediate English levels based on their TOEIC scores. While participating in the English composition class, the participants were enrolled in other English courses in listening, reading, speaking, and Western literature. However, none of these courses involved writing instruction. The participants were 19–21 years old, with an average age of 19.72. Eighteen students, including three higher- and three lower-performing writers from the top and bottom quartiles of each class, were randomly recruited at the end of the program for individual in-depth interviews. The pseudonyms of the higher-performing writers begin with the letter H (e.g., Hana), whereas the pseudonyms of the lower-performing writers begin with L (e.g., Leslie). Ethical concerns regarding interviewing the students are addressed in the data collection and analysis section.

3.2. Instrumentation

3.2.1. AWE tool

Criterion (version. 9.1) was used in this study because most of its local language feedback is instrumental in enhancing EFL writing (Chen et al., 2009) and, between the two relatively more commonly used AWE tools in Asia, most previous studies on AWE effectiveness on Asian EFL learners have examined My Access instead of Criterion, and it appears that none of them has investigated the effects of using Criterion in a multiple-draft approach to assist EFL students in enhancing their grammatical performance.

The main purpose of Criterion is providing students with score reporting and diagnostic feedback on submission of an essay and consequently additional opportunities to practice writing; allowing students to write at their own pace and review, revise, and resubmit their essays; and providing teachers discretionary time to focus on global aspects of student writing and areas of improvement. The topic library of this AWE tool consists of 439 essay topics (each comes with a prompt) of persuasive, expository (including comparison and contrast), descriptive, and narrative writing spanning the academic levels from Grade 4 to college (ETS, 2013).
Criterion identifies 39 error types regarding grammar, usage, mechanics, style, and organization. Users can examine error information in each of the categories by clicking on various tabs near the top of the window (Fig. 1).

3.2.1.1. Feedback function and error-report function. Regarding grammar, the focus of this study, the system identifies nine error types (see Fig. 1 for a detailed list) and provides pop-up notes on the marked errors to explain mistakes. Clicking on the Grammar tab opens a summary error report in a bar graph format that displays the frequencies of various types of grammatical errors identified in a submitted essay. Users can then choose to click on any of the nine error types on the menu bars located to the left of the window to examine a particular type of error that requires their attention.

For example, clicking on “Run-on Sentences” in Fig. 1 opens a new window (Fig. 2) highlighting run-on sentences in the essay. When users roll the cursor over a highlighted error, a pop-up note appears. The pop-up note presents facilitative rather than corrective feedback, and explains the error, provides suggestions that guide users to reconstruct their own texts, and prompts them to use the Writer’s Handbook for further inquiry. The Writer’s Handbook can be accessed by clicking on the link near the upper-right corner of the window shown in Fig. 2, and offers examples of accurate and inaccurate usages for each error type to facilitate feedback interpretation for the evaluated texts. The mechanism assists users in self-editing and directs them to resources where they can reformulate their developing interlanguage (Milton, 2006).

3.2.1.2. Progress-report function. The progress-report function enables users to self-evaluate their performance and progress among various drafts of the assigned essays. Learners can use the information in the progress reports to establish or redefine goals, determine what actions to take next and how, and monitor their subsequent progress. These learning behaviors are indicative of metacognitive strategy use and can potentially facilitate the development of learner autonomy (Gao & Zhang, 2011; Wenden, 2002).

3.2.2. Essay topics
Four comparison essay topics (Appendix A) were selected from the AWE topic library. To ensure that these topics were equally demanding in content, linguistic knowledge, and competence for the participants, first, the researcher administered a 12-item five-point Likert-scale questionnaire (Appendix A) to 128 students from four classes to solicit learner opinions regarding the difficulty level of the topics regarding content development. These students had similar academic backgrounds to the participants of the main part of this study (Section 3.1). The reliability for measuring the difficulty level of each topic was satisfactory, ranging from .88 to .91. The obtained mean scores of 3.41–3.50 indicated that the students considered it moderately easy to develop ideas for the topics. Next, each class was randomly assigned one of the topics. The researcher subsequently identified 12 students in each class; each student’s English proficiency matched that of a student in the other three classes according to their Test of English for International Communication (TOEIC) scores. The essays of these 48 students were separately read by two raters, including the researcher, to identify grammatical errors. The inter-rater reliability was satisfactory at .89. Finally, an analysis of variance (ANOVA) was used to compare the grammatical error frequencies per
Fig. 2. Highlighted texts and pop-up notes describing the problem.
250 words among the essays of the four topics. No significant difference was found ($F(3, 44) = .367, p > .05$), indicating that the topics were linguistically comparable.

### 3.2.3. Questionnaire

A four-scale, 15-item questionnaire (Appendix B) was designed to investigate the participants’ perceived effectiveness of the AWE system (5 items) and their metacognitive strategy use by utilizing the system’s feedback function (4 items), error-report function (3 items), and progress-report function (3 items). The items were measured on a 5-point Likert scale, ranging from 5 (strongly agree) to 1 (strongly disagree). After expert validity and translation reviews of the Chinese version, a trial administration involving three sophomore English majors at a Taiwanese university was conducted to ensure the comprehensibility of the items, after which four items were modified. In the main study, the four scales attained Cronbach’s alpha coefficients ranging from .81 to .89, indicating satisfactory reliability (Table 3).

### 3.2.4. Semi-structured interviews

A 12-question interview protocol was developed according to the suggestions of Patton (2002) to explore the participants’ family background, general learning experience, and their feelings toward and experiences with the AWE tool. Follow-up open-ended questions were asked to probe for unexpected responses or to obtain more in-depth answers. Trial interviews involving two sophomores were conducted to determine whether the protocol required modification. Consequently, two questions that elicited information similar to that obtained from other questions were deleted.

### 3.3. Pedagogical treatment

An eclectic pedagogy was employed, incorporating structural, functional, and process approaches to essay writing. The use of this pedagogy was based on the belief that different writing pedagogies provide complementary teaching routes, with each positioning L2 writing instruction with a specific focus based on the distinctive characteristics of the target learners and learning context (Hyland, 2003; Min, 2009; Richards & Rodgers, 2001). The current study focused on the grammatical aspects of writing; however, considering that EFL learners at basic and intermediate proficiency levels must learn both writing and the English language, the author of the current study incorporated both local and global writing features into the lessons to address the learning needs of the participants. Consequently, structural (focusing on the grammatical aspects), functional (focusing on the organization of comparison essays), and process (focusing on writing processes by using feedback and revising) approaches were adopted. AWE technology was used by students to enhance their grammatical subskills, and writer-teacher conferences involving social interaction and audience awareness development (Chen & Cheng, 2008) were conducted to improve the content and organization of student essays.

<table>
<thead>
<tr>
<th>Essay</th>
<th>Original draft</th>
<th>Revised draft</th>
<th>Paired-samples t tests</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>Essay 1</td>
<td>6.88</td>
<td>7.82</td>
<td>3.83</td>
</tr>
<tr>
<td>Essay 2</td>
<td>5.08</td>
<td>6.97</td>
<td>2.17</td>
</tr>
<tr>
<td>Essay 4</td>
<td>1.71</td>
<td>1.57</td>
<td>.80</td>
</tr>
</tbody>
</table>

Note. * per 250 words; **p < .01.

<table>
<thead>
<tr>
<th>Comparison of original drafts</th>
<th>Essay A</th>
<th>Essay B</th>
<th>t</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Essay 1 – Essay 2</td>
<td>6.88</td>
<td>7.82</td>
<td>5.08</td>
<td>6.97</td>
<td>1.59</td>
</tr>
<tr>
<td>Essay 2 – Essay 3</td>
<td>5.08</td>
<td>6.97</td>
<td>3.42</td>
<td>2.52</td>
<td>1.83</td>
</tr>
<tr>
<td>Essay 3 – Essay 4</td>
<td>3.42</td>
<td>2.52</td>
<td>1.71</td>
<td>1.57</td>
<td>5.39</td>
</tr>
</tbody>
</table>

Note. * per 250 words; *p < .05, **p < .01.
In Weeks 1 and 2, the participants were instructed how to use the point-by-point and block methods to organize comparison essays, and they were exposed to various writing strategies through analyzing sample essays and practicing outlining (functional approach). In Weeks 3–9, the students composed four assigned comparison essays that comprised multiple drafts (process approach), using the Criterion website in class. No minimal word count was required; however, it was suggested that the students compose essays of four to five paragraphs, including an introductory paragraph, a two- or three-paragraph body, and a concluding paragraph. The week after writing an original draft, the students were required to revise their essay in class based on the grammatical feedback of the AWE system (structural and process approach). One-on-one and small-group (two to four students) teacher-writer conferences were conducted in- and out-of-class to address concerns about the content and organization of individual student essays (functional and process approaches), and to compensate for the lack of human interaction using AWE alone. All students participated in the conferences three to four times. Although an eclectic pedagogy addressing both local and global issues was used to maximize learning, the current study focused on students’ development of grammatical accuracy, and the content and organizational aspects of the essay are beyond the scope of this study. To ensure that only the pedagogical treatment of Criterion caused the obtained results in grammatical accuracy, the teacher—writer conferences strictly focused on discussing the global aspects of writing. If students raised grammar questions, they were directed to revisit the AWE feedback portal and independently seek answers by using Criterion resources. A pre-study conference was conducted between the researcher and the instructors to ensure that the instructors were familiar with the procedure and the importance of maintaining the treatment integrity of the research by following the lesson plan protocol.

In addition, based on the guiding principle of Milton (2006), the instructor in each of the participating classes demonstrated how students could access online resources, including the Writer’s Handbook provided in this AWE system, if they required additional information pertaining to a specific grammatical point in the future. In addition, the instructor encouraged the participants to use metacognitive strategies (i.e., thinking about learning, planning, self-monitoring, and self-evaluation; Oxford, 1990), by demonstrating how students could utilize the various functions in the system, including the error report and progress report, to reflect on their learning and monitor their writing development. Such modeling was conducted thoroughly at the onset of the writing program and subsequently in the mini-lessons as a scaffolding mechanism three to four times during the program.

Because of administrative constraints, 9 weeks was the maximal length for the classroom implementation portion of the current study. To ensure that all classes received the same treatment, it was determined that Essays 1, 2, and 4 should comprise writing the first and revised drafts, whereas Essay 3 comprised only a first draft. The writing process is depicted in Fig. 3.

### 3.4. Data collection and analysis

The grammatical feedback reports generated by Criterion for the participant essays were collected to address RQ 1. The questionnaire and interviews were administered to answer RQs 2 and 3. When recruiting participants for the interviews, the researcher informed all the students from the top and bottom quartiles of each class about the research objective and interview process, and assured them that their rights and privacy would be protected, including shielding their identities from their instructors, during the data collection, analysis, and report stages. An incentive of approximately US$7 was offered to each interviewee to compensate for their time. Among the 22 volunteers, nine from the top quartile and nine from the bottom quartile were randomly selected for the interviews, which were conducted by two of the writing instructors who had received training and were experienced in conducting interviews. To avoid flawed data because of power hierarchy or over-identification with the interviewer (Glesne, 2010), the participants were not interviewed by their own instructor. Consent forms and verbal explanations were utilized to reassure the participants of privacy protection before the interviews. Each interview took approximately 45 min at an on-campus location selected by each interviewee. Because of the proficiency level of the participants, the interviews were conducted in Chinese to facilitate communication. The voice-recorded interview data were immediately assigned pseudonyms to protect the identities of the participants. The interviews were transcribed verbatim and presented to the participants for verification. The transcripts were translated into English by the researcher and subsequently examined by two proficient bilingual speakers to minimize the loss of the original meanings.
Although the essay topics were comparable (see Section 3.2.2), it seemed reasonable that the participants composed longer essays (i.e., an average of 210, 213, 232, and 292 words in the original drafts of the four essays) when they gradually became more familiar with structuring and expressing ideas in a comparison essay. Therefore, the data were analyzed according to the average frequencies of errors per 250 words in each draft. Repeated errors were counted repeatedly. Paired-samples t tests and one-way repeated measures ANOVAs and subsequent protected t tests, a post-hoc analysis for within-subjects factors (Howell, 2013), were used to address RQ 1; descriptive statistics and Pearson correlations were used to address RQ 2.

The data analysis for RQ 3 was conducted in two stages (Chao, 2015). First, employing a constant comparative method (Miles & Huberman, 1994), the researcher and her colleague read through the 18 interview transcripts to identify meaningful statements and categorize recurrent themes related to the exercise of learner agency through repeated discussion, comparison, and contrasting. Open coding was used to identify distinct categories in the data, and axial coding was used to explore the relationships among the categories through a recursive coding and recoding process. Redundant categories were deleted, and similar categories were combined. Subsequently, selective coding was adopted to examine the saturation of categories. When the recursive open, axial, and selective coding processes were completed and the researcher and her colleague agreed that all the categories were assigned meaningfully and no new category emerged from the data, four general types of learner profiles related to the exercise of learner agency were determined. Participants who were strongly goal-oriented and self-motivated in learning were identified as goal getters (5 cases). Those particularly concerned about accuracy in their language output were labeled as accuracy pursuers (4 cases). Participants who seemed to lack learning motivation and tended to learn passively were identified as reluctant learners (3 cases). Finally, the participants who did not become aware of their learning until the later part of the writing program were categorized as late bloomers (6 cases). When a learner displayed more than one type of characteristics, the categorization was based on his or her most salient characteristics. Next, the interview data of one participant from each of these four types were selected for reanalysis following the qualitative analysis procedure of coding, categorization, description, and interpretation proposed by Patton (2002) to address RQ 3. To ensure trustworthiness, the participants were asked to verify the analytical interpretations developed by the researcher.

4. Results

4.1. RQ 1: At what point in the AWE-assisted process-writing program does learners’ grammatical performance change?

The paired-samples t tests (Table 1) comparing the grammatical error frequencies of the original and revised drafts of each essay (i.e., Essays 1, 2, and 4) revealed significant differences between all comparison pairs, indicating that students progressed grammatically in each of the revised essays.

An ANOVA was used to compare the means of the grammatical error frequencies in the original drafts of the four essays. Because a significant difference was observed ($F(1, 62) = 95.608, p < .0001$), subsequent protected t tests were conducted to pinpoint the differences. Table 2 shows no significant difference between the original texts of Essays 1 and 2. However, significant improvement was detected between the original texts of Essays 2 and 3, and between Essays 3 and 4. This indicates that significant improvement in new texts was not observed until the original draft of the third essay.

4.2. RQ 2: How does the grammatical performance as evaluated by Criterion relate to learner perceptions of the effectiveness of the system and learners’ self-reported metacognitive strategy use?

To address this research question, learner perceptions were first examined using the questionnaire. As indicated in Table 3, the participants generally considered Criterion instrumental in enhancing their writing.

The learner perceptions and experiences revealed by analyzing the interview data supported the statistical results. All 18 informants indicated that the AWE feedback identified their grammatical errors. However, among these 18 informants, it seems that Criterion provided scaffolding that was more effective for higher-performing learners than for lower-performing learners: Seven of the nine lower-performing writers stated that they were occasionally unable to comprehend the error messages, whereas none of the higher-performing informants appeared to encounter such a problem. All nine lower-performing writers and one of the nine higher-performing writers mentioned that they occasionally experienced difficulty in revising their texts by referencing the auxiliary resources in the Writer’s Handbook.

The questionnaire was further used to investigate the learners’ metacognitive strategy use facilitated by various AWE functions. As shown in Table 3, the learners reported their use of metacognitive strategies facilitated by the Criterion feedback function; however, their self-reported learning behaviors indicated that the error- and progress-report functions of Criterion were less effective in inducing their metacognitive strategy use.

To examine how grammatical performance as evaluated by Criterion related to learner perceptions and metacognitive strategy use, Pearson correlations were conducted. Grammatical performance was calculated based on the number of errors made, which were denoted using a minus sign (e.g., three errors were recorded as “−3”). The correlation matrix in Table 4 indicates significant and positive relationships among all of the variables, including learner perceived effectiveness, their self-reported metacognitive strategy use utilizing the AWE feedback, error-report, and progress-report functions, and grammatical accuracy in both the original and revised drafts of Essay 4.
4.3. RQ 3: Based on learner narratives, what additional factors mediate learner grammatical development in this AWE-assisted writing program?

RQ 3 was addressed using the interview data. As discussed previously, four general types of learner profiles and experiences related to the exercise of learner agency were identified among the 18 informants in the first phase of the data analysis. Subsequently, the narratives of one student from each category were reanalyzed. These four students were selected because their narratives were the most salient and described incidents and properties that best reflected each respective category. They were (a) Helen, a goal getter; (b) Hana, an accuracy pursuer; (c) Lori, a reluctant learner; and (d) Leslie, a late bloomer. Although these participants’ stories were unique and not to be generalized to other learners who were not interviewed, their experiences are insightful because they reveal EFL students’ learning processes in the context of an AWE-assisted writing program. Learner agency, learning style, metacognitive strategy use, language gap noticing, and process writing appeared to mediate the AWE-assisted learning of these informants, albeit in different ways.

4.3.1. Helen: a goal getter

Helen was a self-motivated language learner who clearly knew what she wanted to accomplish. To realize her learning and professional goals, she worked diligently, particularly in classes that she considered instrumental to achieving her translation professional goal, including the writing course of this study. Growing up in a middle-class family with highly educated parents who greatly valued her education, Helen’s strong goal orientation seemed to be affected by the way she had been raised. Although her parents had participated actively in Helen’s learning process, they had rarely given her precise instructions; instead, they provided general directions and encouraged her to pursue improvement through constant self-monitoring. The facilitative role of her parents perhaps had helped train Helen to consider and manage learning in a reflective manner. In the writing class, the AWE functions helped Helen self-evaluate her learning status and self-monitor her grammatical progress:

What I loved most about Criterion was, after I submitted my essay, the trait feedback analysis immediately graphically presented my errors by error types. That way I could easily identify my weakest grammar points at a glance. I think understanding my own strength and weakness is very important… Before I began to draft a new essay, I reminded myself to read the error feedback from the previous essays so that I would not make the same mistakes again. It might not make me error free, but at least it helped me reduce the number of errors, and making progress is important to me.

This interview data reveals three possible mediators of Helen’s writing development in this AWE-assisted process-writing program. The first is agency. Helen was apparently self-disciplined and highly aware of what was crucial in her writing development. Possibly influenced by the way her parents guided her learning, she believed that identifying her strengths and weaknesses during the writing process and subsequently making progress on the basis of that understanding were critical. The second possible mediator is learning style. Helen mentioned twice during the interview that she could easily evaluate her grammatical performance because the AWE system depicted a graphical overview of her writing strengths and weaknesses. This projected Helen’s visual learning style, which enabled her to spontaneously exploit the graphical presentations as a tool for strategic learning.

The third possible mediator is metacognition. Goal setting (making progress), self-evaluation (identifying current grammatical performance), and self-monitoring (using newly developed self-evaluation and newly acquired linguistic knowledge from error feedback to manage subsequent learning) seemed to have assisted Helen in recognizing her language gap, leading her in a positive direction of writing development. Helen described the gap between her interlanguage and target language form as “unbeatable cockroaches” that she vowed to conquer:

The trait feedback analysis helped me identify ill-formed verbs as my major problem area the first time I used Criterion. Since then, I paid attention when I used a verb and I used the error reports and progress reports to ensure that I continued improving in this regard. Although the ill-formed verbs were like unbeatable cockroaches contaminating the quality of my essays, at least I managed to keep the number down, and I will continue to eliminate them. If the teacher...
had required us to write only one draft per essay, and if we had not used Criterion to check our grammar each time, I would not have been able to notice my progress in reducing ill-formed verbs.

It appears that the multiple-draft approach raised Helen’s awareness of her language gap, enabling her to consciously monitor her effort and progress in narrowing the gap.

4.3.2. Hana: an accuracy pursuer

Similar to Helen, Hana also grew up in a family that valued education. However, in contrast to higher-order thinking, which was strongly encouraged by Helen’s parents, precision appeared to be highly emphasized in Hana’s family. Hana recalled that she was required to fold each pair of socks in exactly the same way, sort her T-shirts according to colors, and keep her handwriting neat and error free. Such a parental style may have somehow shaped Hana’s belief in writing accuracy:

My previous writing teacher told me not to care too much about grammar. She was probably right asking us to focus on meaning, but (sighs) I wanted to do things right. Knowing that there were grammar mistakes in my writing made me uncomfortable (sighs). I’m a perfectionist. My parents taught me to be precise. Besides, my dream is to become a famous cram-school English teacher. How can I become one if my English is not perfect? … I felt excited to use the writing software this semester. It identified my errors immediately. I used the error and progress reports to track my errors and ensure that I continued improving. How magical! I guess neither my current nor previous writing teachers can be this super powerful [providing detailed feedback instantly], especially, as they mentioned, they wanted to spend more time on the contents and meaning of our writing... When I received [AWE] feedback, I took note of the errors to avoid making the same mistakes in the future. It is important to write correctly. Each time I saw the number of my grammar errors reduce in the software reports, I felt a sense of satisfaction and I smiled or even cheered for myself (laughs).

From these narratives, agency and metacognition appeared to mediate Hana’s writing development in the AWE-assisted process-writing program. First, Hana’s frustration in receiving insufficient corrective feedback in her previous writing classes reflected the expectations of typical students from authoritarian cultures in desiring corrective feedback (Hyland & Hyland, 2006). The “doing-things-right” perfectionist characteristic shaped by her family education appeared to have reinforced Hana’s pursuit of precision in writing. The goal of becoming a star teacher might have further strengthened her seeking of perfection in grammatical accuracy. These manifestations of agency could have jointly led Hana to assign relevance and significance to the AWE instant feedback and stimulated her metacognition in goal setting (writing correctly), self-monitoring (purposeful use of the note-taking strategy to enhance the accuracy in subsequent essays), and self-evaluation (using the AWE reports to self-evaluate her performance).

4.3.3. Lori: a reluctant learner

Agency also appeared to mediate Lori’s learning and performance, albeit differently from how it affected Hana and Helen. According to Lori, she wanted to study fashion design but was “forced” by her father to study English. Now that she was a sophomore English major, she still perceived herself as “a loser with no talent in English learning” and described that she was “a fish out of water” because she lacked interest in the major. Lori was determined to pursue her dream in fashion design: “I’ll earn this degree for my dad. But after I’ve got the piece of paper, I’ll do what I want, not what he wants.” Likely because of such a passive attitude toward her academic major, in contrast to Helen and Hana intentionally employing metacognitive strategies by using various AWE functions to enhance their learning, Lori did not seem to employ any metacognitive strategy. Lori stated, “I revised my essay based on the feedback, but I didn’t bother to remember the mistakes... We were asked to revise and resubmit the essay. That was what I did.”

In addition to agency, learning style might have mediated Lori’s learning and performance, although negatively. According to Lori’s narratives, she learned more happily and productively when it involved interacting with people or being physically active, such as participating in interviews, group presentations, drama performances, and games. By contrast, she felt frustrated reading and interpreting the AWE comments without human interactions. Lori’s social and kinesthetic learning tendency appeared to hinder her grammatical development in this AWE-assisted program, in which being reflective of one’s own learning played a crucial role.

4.3.4. Leslie: a late bloomer

Like Lori, Leslie became an English major according to her parents’ wishes. However, although Leslie did not seem to have a strong passion for learning English, in contrast to Lori, she was not resistant. From her narratives emerged an image of a family with authoritarian parents and an obedient daughter. Leslie’s parents made almost every decision regarding her learning, including scheduling her study hours, determining the types and levels of English proficiency tests she should take, and selecting student clubs and elective courses. Leslie appeared to feel comfortable letting her parents make such arrangements for her, whether trivial or important. “I believe they make right decisions for me,” stated Leslie. Possibly because of this obedient, “go-with-the-flow” characteristic, reflection did not occur in Leslie until after six drafts of essay writing:

In the first few weeks, Criterion gave me instant feedback and I revised the errors accordingly. When I received the feedback for the first draft of Media Portrayals of Teenagers [the fourth essay], it suddenly hit me. My weakest spot in grammar was the use of because! I used a period at the end of dependent clauses when I wrote the earlier essays. Although I corrected the punctuation each time the trait feedback analysis indicated it was a fragment, I was like a
machine executing an order without realizing I was continually making the same mistakes... I didn't know why it took me so long to become aware of such an obvious weakness in my writing, but I knew from that point on I had to pay attention when I wrote a sentence starting with because.

Although using Criterion feedback did not initially activate metacognitive strategies, the repeated use of the system ultimately helped Leslie realize the problem area requiring the most attention (i.e., self-evaluation), and prompted goal setting, indicating the process-writing approach as a possible mediator of gap noticing and metacognitive behaviors.

5. Discussion

This study evaluated learners’ grammatical performance in revised and subsequent essays in a process-writing program assisted by AWE technology. Although enhanced grammatical accuracy was observed in each essay revision, new texts showed no improvement until the composition of the third essay. The differential effects between enhancing the grammatical accuracy of revisions and new texts are consistent with skill acquisition theory (DeKeyser, 2007). After the AWE feedback system explicitly conditioned the students to initially interpret and become aware of English grammatical rules (i.e., presentation of declarative knowledge), the recursive multi-draft process offered the learners opportunities to internalize the diagnostic feedback by revising texts (i.e., practice of procedural skills), as evidenced by the improvement in the revised essays. The integration of procedural skills in turn led to gradual automatization and longer-term improvement, as evidenced by the decreased number of grammatical errors in the original drafts of the last two essays. The proceduralization of declarative knowledge required less practice from the participants, compared with that required before a specific language skill could become automatic in producing new texts.

The positive outcomes in this study are consistent with previous research that has also employed autonomous computer-aided writing pedagogies that have enhanced learner responsibility and relieved the burden and tedium of writing instructors’ having to respond repetitively to local concerns. Such positive results can be discussed pertaining to reactive autonomy (Littlewood, 1999) in the assessment-for-learning paradigm. After the participants were introduced to the learning direction and notion of learner autonomy, they participated in identifying and addressing problems by using AWE feedback, and self-evaluated their weaknesses, strengths, and progress by using error and progress reports. When the participants perceived that the feedback was insufficient and could not respond to the identified problems, they could self-access the Writer’s Handbook and online resources to address these problems. It is reasonable to postulate that each of these learning behaviors helped the learners move toward reactive autonomy. Thus, during the cyclical process of composing and revising, the role of the AWE system was more a learning facilitator than that of an end-product assessor (Dikli, 2010; Stevenson & Phakiti, 2014). In line with Wang and Goodman (2012), the formative characteristic of Criterion was emphasized when the AWE system was integrated into the process-writing pedagogical design, which in turn likely facilitated learner responsibility for writing and learning (as demonstrated by Leslie’s determination to produce correct dependent clauses instead of passively making corrections after receiving error messages), and activated learning management and cognitive processes (as demonstrated by Hana using AWE progress reports and taking notes to track her improvement), both of which are essential to autonomy (Gao & Lamb, 2011; Gao & Zhang, 2011; Littlewood, 1999).

Both quantitative and qualitative inquiries indicated positive learner perceptions of the AWE system in facilitating grammatical development. However, seven of nine lower-performing informants’ occasionally experiencing difficulty understanding and addressing the feedback suggests that teacher or peer scaffolding beyond AWE assistance might be necessary to facilitate productive learning for lower-performing writers (Dikli, 2010; Wang et al., 2013).

A strong positive statistical correlation was observed between writing performance and self-reported metacognitive strategy use; learners who reported using the feedback, error-report, and progress-report functions of the AWE system to generate a metacognitive view of their current grammatical performance and plan for subsequent learning tended to demonstrate a higher level of grammatical accuracy at the end of the writing program (Table 4). These statistics were supported by the interview data. Among the 18 interviewed participants, higher-performing writers reported employing metacognitive strategies more than their lower-performing peers did. For example, as shown in the previous excerpts, Helen used the graphical representations in Criterion for self-evaluation, and Hana consciously attempted to avoid making the same linguistic errors. By contrast, among the lower-performing informants, Leslie’s metacognition was not active until the final essay, and Lori simply corrected the identified errors without reflection. These findings correspond with the language learning strategy theory of Oxford (1990), that metacognitive aids assist in focusing learner attention to recurring errors; they are also consistent with previous empirical studies regarding the positive effects of metacognition on language learning and writing skill development (Rotterall & Murray, 2009; Schoonen, van Gelderen, Stole, Hulstijn, & de Glopper, 2011).

In agreement with Ferris (2006) and van Lier (2008), learner agency and learning styles emerged from the interview narratives seemingly mediated the participants’ writing development in various ways in this AWE-assisted writing program. For instance, feeling forced by her father to major in English appeared to make Lori a reluctant learner paying minimal attention to the AWE feedback, whereas the perfectionist characteristic of Hana seemingly led her to exploit the various AWE functions to the full extent. Helen’s visual learning style appeared to facilitate her analysis of personal strength and weakness through her use of the graphical presentations of the AWE grammatical reports, which is in contrast to the group and kinesthetic learning styles that appeared to discourage Lori’s learning in an AWE program, which required individual work and reflection.
Among the 18 interviewed participants, four learner profile types related to the exercise of learner agency were deduced: goal getters (5 cases), accuracy pursuers (4 cases), reluctant learners (3 cases), and late bloomers (6 cases). The repeated act of gap noticing and metacognitive strategy use mediated by process writing appeared to facilitate the writing development of all learner types except the reluctant learner category. Automated evaluation provided prompt feedback and enabled the instructors in this study to implement the process-writing approach effectively with multiple rounds of writing and revisions. Reflecting Schmidt’s (2012) noticing hypothesis and Ortega’s (2009) interlanguage theory, the process-writing approach implemented in the present study appeared to generate both initial opportunities for the student writers to consciously discern the discrepancies between their interlanguage output and target language form, and additional opportunities to practice and process the linguistic forms sufficiently for longer-term retention. Process writing also seemed to mediate learners of distinct profile types to use various modes of metacognitive strategies, including goal setting (Helen, a goal getter; Hana, an accuracy pursuer; and Leslie, a late bloomer), planning and self-monitoring (Helen and Hana), and self-evaluation (Helen, Hana, and Leslie). However, the activation of metacognition in the goal getters and accuracy pursuers appeared to occur earlier than in the late bloomers in the writing process. Without the multiple opportunities generated by the process approach, late bloomers such as Leslie might not notice their learning gaps or process the recognition with sufficient depth to activate goal setting and subsequent learning behaviors. Based on these findings, elements were added to the writing process shown in Fig. 3, resulting in a recursive composing process that involves agency, learning styles, noticing language gaps, using metacognition, and undergoing three skill acquisition stages, as shown in Fig. 4.

By employing a multiple-draft process-writing approach integrating the use of AWE technology, the EFL novice writers in this study appeared capable of modifying the grammar of their revisions based on the Criterion feedback and, subsequently, applying the acquired skills in composing new texts. Although the accuracy rate of Criterion linguistic feedback is not yet ideal based on several system-centric evaluations (Chen et al., 2009; Chodorow et al., 2010; Dikli & Bleyle, 2014), Chodorow et al. (2010) indicated that the overall linguistic accuracy in a text increases if a learner accepts all AWE suggestions, and that AWE could achieve a higher level of performance in the future with continued technological advancement. In contexts such as many parts of Asia, where teachers are highly respected, the level of learner take up from teacher feedback could be high. However, similar to machine feedback, teacher feedback on form might not always be accurate. As Chen et al. (2009) argued, when EFL writing instructors are non-native speakers of English, some feedback on form might be inaccurate. Moreover, Dikli and Bleyle (2014) argued that it is impractical for a teacher to provide detailed feedback on multiple essay drafts regularly. Assuming that the class enrolment in Taiwan and elsewhere remains large, both writing teachers and students could benefit from an AWE system that provides timely feedback, enables frequent writing practice and opportunities to notice language gaps and strengthen the proceduralization of new knowledge, fosters learner autonomy, and reduces teacher workloads in addressing surface errors (Chen et al., 2009). Freeing up time by employing an effective AWE system enables EFL writing teachers to concentrate on providing feedback in global language domains such as content and discourse structures (Lai, 2010), thus benefiting learners in developing the critical-thinking aspects of their writing.

6. Conclusions

The current study was conducted on the basis of the notion that an effective AWE system pertaining to local aspects of writing could help relieve teacher workload of the tedious and often repeated tasks of responding to grammatical errors, and consequently free up their time to focus on global writing development. This study used a process pedagogy incorporating the use of Criterion feedback to mediate learner consciousness of language gaps, thereby causing subsequent linguistic restructuring and enhancing learner grammatical performance in both revisions and subsequent new texts. Likely because integrating and internalizing knowledge and skills is gradual and incremental (DeKeyser, 2007), and multiple opportunities for noticing gaps and practices are required for deep processing and internalizing noticed language forms (Ortega, 2009; Schmidt, 1990; Tode, 2008), students’ grammatical performance first improved in their revisions and later new writing. The positive performance seemed to be facilitated by repetitive practices and gap noticing, which were in turn facilitated using the AWE system under the integrated process and structural pedagogy.

The findings yield several pedagogical implications. First, it appears that lower-performing writers such as those in this study require additional scaffolding to comprehend Criterion feedback and resources to address the linguistic problems identified by the system. Therefore, human scaffolding from teachers or from peers who are more advanced should be provided in addition to machine assistance to ensure successful learning. Second, Criterion allows writing teachers to reduce the amount of time spent on the local concerns of student writing; this time could be used effectively by helping students develop global composition skills (Lai, 2010). Third, because of the immediacy of the AWE diagnostic feedback, teachers can increase the number of writing assignments to build student ability and self-efficacy in written English communication. Fourth, such an AWE system has an optimal effect when learners use it for multiple rounds of essay writing because, as Schmidt (1990) and Tode (2008) have asserted, noticing facilitates only preliminary registration; additional opportunities for noticing must be provided for learners to process the language forms deeply enough to facilitate retention. This might particularly apply to learners at lower proficiency levels or those lacking metacognitive knowledge and experience.

Finally, teachers should provide appropriate guidance to help students develop reactive autonomy by independently planning and executing writing tasks. Writing teachers should inform students of the rationale for using AWE at the onset of writing programs, explicitly explain the expected change in the learner role, and explicate that autonomy is required for and indispensable to becoming an effective writer. To ensure productive autonomous learning, students should be taught...
explicitly how and when to use online resources to reformulate their own developing interlanguage (Milton, 2006). The advantages of metacognition should be clearly explained, and methods for facilitating metacognition by using the functions of the AWE system should be demonstrated before the system is employed.

This study has certain limitations that future studies could address. First, because of administrative constraints, the third essay in this study involved original but not revised texts. A future longitudinal study could avoid this type of research limitation by observing the long-term effects of employing Criterion. Second, this study focused on quantitative and retrospective inquiries of the effects of Criterion. Introspective methods, such as thinking aloud, can be employed in future studies to elucidate the inner cognitive processes of EFL learners and explore deeply the differences in metacognitive strategy use among students of different learning profiles when using the AWE system. Third, without a control group, the AWE-assisted process approach could not be isolated as the sole attributer of the positive results observed in this study. Potential enhanced familiarity with comparison essay structures and idea development likely reduced the participants’ cognitive loading and allowed them to focus more on the grammatical aspect of writing, thus reducing the number of grammatical errors. In addition, the students’ progress could simply be a result of maturation or, in other words, a function of time (Gravetter & Wallnau, 2004). Therefore, future research should include a control group to ascertain the causal relationship between using an AWE-assisted
process approach and grammatical learning outcomes. Finally, the present study examined learner performance by holistically considering various error types. Future studies can further examine the nature of error frequency reduction in distinct error types by adopting both quantitative inferential and qualitative in-depth case-study research designs.

Acknowledgments

The author is grateful to Drs. Wai Meng Chan, Hsin-I Chen, and the editors and reviewers for their valuable insights and suggestions regarding this manuscript.

Appendix A. English essay topics questionnaire.

INSTRUCTIONS: Please circle the number that best describes your views on and experiences with using Criterion to improve the linguistic aspects of your writing. There are no “right” or “wrong” answers. Your opinion matters.

Different Teaching Styles
Different teachers might have unique teaching styles. In this essay, you are asked to compare the teaching styles of two of your teachers, including different or similar philosophies and approaches.
1. It is easy for me to develop ideas for this topic.
2. I have trouble thinking about what to write on this topic.
3. It is easy for me to come up with the content for this essay.

Fictional Character and Me
In this essay, think about a fictional character from an entertainment medium (e.g., a novel, fairy tale, movie, situation comedy) that is similar to you in some way and compare the character with you.
4. It is easy for me to develop ideas for this topic.
5. I have trouble thinking about what to write on this topic.
6. It is easy for me to come up with the content for this essay.

Me in the Eyes of Family and Friends
People sometimes act one way around their family and another way around their friends. In this essay, you are asked to compare what your family and friends think you are like.
7. It is easy for me to develop ideas for this topic.
8. I have trouble thinking about what to write on this topic.
9. It is easy for me to come up with the content for this essay.

Teenagers in the Media and in Reality
In this essay, you are asked to think about the teenage characters in your favorite movies or television shows and compare the portrayals of teenagers in media with those in real life.
10. It is easy for me to develop ideas for this topic.
11. I have trouble thinking about what to write on this topic.
12. It is easy for me to come up with the content for this essay.

Appendix B. Criterion satisfaction & writing strategy use questionnaire. [Grammatical accuracy]

INSTRUCTIONS: Please circle the number that best describes your views on and experiences with using Criterion to improve the linguistic aspects of your writing. There are no “right” or “wrong” answers. Your opinion matters.
1. Using the feedback function helped me understand my writing performance.
2. It was easy for me to understand the feedback.
3. The feedback identified the problems in my writing.
4. I revised my essays based on the feedback.
5. Criterion provided comments that helped improve my grammar.
6. I read the feedback carefully to remember my mistakes for future improvement.
7. I referred to Criterion feedback for my previous essays to avoid making the same errors again.
8. I read the error report after an essay was submitted.
9. I used the error report to identify the problem areas in my writing.
10. I used the error report to analyze my writing problems.
11. I read the progress report after an essay was submitted.
12. I used the progress report to understand how I was progressing.
13. I used the progress report to analyze my writing strengths and weaknesses.
14. I used the trait feedback analysis on my grammar to identify the areas that required the most attention in my writing.
15. I used the trait feedback analysis on my grammar to analyze my writing problems.

References