Enhancing the use of evidence in argumentative writing through collaborative processing of content-based automated writing evaluation feedback

Zhan Shi, The University of Hong Kong
Fengkai Liu, City University of Hong Kong
Chun Lai, The University of Hong Kong
Tan Jin, South China Normal University

Abstract

Automated Writing Evaluation (AWE) systems have been found to enhance the accuracy, readability, and cohesion of writing responses (Stevenson & Phakiti, 2019). Previous research indicates that individual learners may have difficulty utilizing content-based AWE feedback and collaborative processing of feedback might help to cope with this challenge (Elabdali, 2021; Wang et al., 2020). However, how learners might collaboratively process content-based AWE feedback remains an open question. This study intends to fill this gap by following a group of five Chinese undergraduate EFL students’ collaborative processing of content-based AWE feedback on the use of evidence in L2 argumentative writing during five writing tasks over a semester. Student collaboration was examined through tracking the recordings of collaborative discussion sessions as well as their written drafts and revisions, and through collecting interview responses from individual learners. The findings revealed that the collaborative processing of AWE feedback was experienced in three phases, namely the trustful phase, skeptical phase, and critical phase. Although content-based AWE feedback could facilitate the development of some aspects of evidence use, collaborative discourses were instrumental in developing learners’ understanding and skills for certain aspects of evidence use that AWE feedback failed to address. The findings suggest that collaborative discourse around content-based AWE feedback can be an important pedagogical move in realizing the potential of AWE feedback for writing development.

Keywords: Content-based AWE Feedback, Collaborative Processing of Feedback, Using Evidence in Argumentative Writing

Language(s) Learned in This Study: English


Introduction

In recent decades, the use of Automated Writing Evaluation (AWE) has become increasingly pervasive in a variety of educational contexts. It is acclaimed for providing evaluative feedback for writers and freeing up teachers’ time in writing instruction (Stevenson, 2016). A burgeoning number of AWE systems for ESL writing have been developed, such as Criterion, Grammarly and so on, that can generate corrective feedback regarding global organization, language use, and mechanics in students’ writing (Deane, 2013). These AWE systems, albeit effective in providing written corrective feedback, mainly concentrate on language-based error correction and are unable to address the complex aspects of writing (Vojak et al., 2011). Recently, AWE systems that provide feedback on content-based features have started to emerge,
such as *EssayCritic*, *eRevise*, and *Virtual Writing Tutor*. These systems can evaluate the quality of essays, argumentative writing in particular, a genre that is prominent in international language testing and academic contexts (Lee, 2020). These AWE systems can offer immediate and individual-specific feedback on features at the content level, and have been found to yield positive results on students’ practice of revision (Mørch et al., 2017). Despite the potentials of AWE systems, students have been found to struggle with using content-based AWE feedback to make effective revision as they might lack sufficient revision strategies to address the AWE feedback (Wang et al., 2020). Thus, even though AWE systems help learners identify their problems in writing, external support is needed to enhance learners’ revision skills to achieve substantive improvement at the content level (Li et al., 2015).

Collaborative processing of AWE feedback might be a plausible solution to cope with the above challenge, as making sense of the feedback, devising revision plans, and making revisions in response to the feedback together might promote deeper engagement with feedback and enhance learning (Kim & Emeliyanova, 2019; Wigglesworth & Storch, 2012). However, limited research has explored collaborative processing of content-based AWE feedback. As collaborative processing of feedback might enhance its efficacy, it is vital to probe into the collaboration around content-based AWE feedback so as to shed light on potential pedagogical approaches to maximize the potentials of content-based AWE. This study adopted a case study design to unveil how a group of intermediate university English language learners collaboratively processed feedback from a content-based AWE (*Virtual Writing Tutor*) on their argumentative writing in an English for Academic Purposes (EAP) class in China.

**Literature Review**

**Content-based AWE Systems**

Content-based AWE systems are an emerging innovation for enhancing the efficacy of AWE for writing development. The first few content-based AWE systems designed to evaluate the quality of argumentative writing primarily focused on the presence of themes in the arguments. The *EssayCritic* system, for instance, can detect the presence and absence of *sub-themes* through identifying the specific concepts, topics, and ideas in the essay (Mørch et al., 2017). When a missing sub-theme is identified, *EssayCritic* will instantly generate feedback to prompt students to include more content in the essays. Mørch et al.’s (2017) study has shown that such content-based AWE feedback could trigger students to focus more on content creation in the process of revision.

Recently, content-based AWE systems have expanded to provide feedback on the use of evidence, a key component in argumentative writing. The use of evidence requires learners to draw upon reliable sources of information to validate an argument (Kibler & Hardigree, 2017). This practice usually consists of the incorporation of evidence sources, the elaboration of the relevant details, and the construction of a sound reasoning process (Jin et al., 2019). Albeit important, being able to use evidence effectively is often viewed as a challenge in writing (De La Paz et al., 2012). This challenge partially derives from the insufficient provision of formative feedback on the use of evidence in the writing classes (Mitchell & Andrews, 2000), as it is time-consuming for teachers to render feedback on a large scale. In this regard, leveraging AWE systems becomes a plausible intervention to enhance learners’ use of evidence.

Given the critical role of using evidence in argumentative writing, some researchers have developed AWE system that incorporated the use of evidence as a content-based feature. For example, Wang et al. (2020) have developed *eRevise*, an AWE system that can evaluate learners’ use of text evidence in response-to-text assessment. The researchers found that the majority of students had made an attempt to revise their use of evidence based on the AWE feedback generated by *eRevise*. Similarly, *Virtual Writing Tutor*, an open-access AWE system that can evaluate both form-focused features (vocabulary, grammar, spelling etc.) and content-based features (such as argument structure and evidence use), has also been found effective in improving undergraduate EFL students’ academic writing skills (Al Badi et al., 2020).
Despite the accumulating evidence on the potential of content-based AWE as a formative assessment tool for writing development, content-based AWE systems such as eRevise may fail to recognize the nuances in students’ revisions (Wang et al., 2020). Such a limitation resonates with scholars’ reservations about the alleged capabilities of AWE systems to promote writing development (Hegelheimer & Lee, 2013). The reservations partly derive from the contentious epistemology of writing that underpins these programs. AWE systems with built-in linguistic norms tend to view writing as an accuracy-focused rather than socially situated activity and are likely to “reinforce narrow conceptualizations of good writing as being the equivalent of grammatically correct writing, without considering other important aspects, such as development of ideas, coherence and creativity” (Stevenson & Phakiti, 2019, p. 136). This epistemology may have a dehumanizing effect since “the one-size-fits-all nature of AWCF…takes little or no account of individual differences” (Ranalli, 2018, p. 654).

Scholars also question the usefulness of content-based AWE feedback because of its lack of explicitness, as many content-based AWE programs only provide indirect feedback and inadequate concrete suggestions to guide writers’ remedial action, which may render it a particular challenge for learners with low proficiency levels (Ranalli, 2018). As acknowledged by Lee (2020), even though a content-based AWE feedback message may prompt students to explore and include more suggested ideas, “the extent to which the suggested ideas could be elaborated is at individual student’s discretion” (p. 46). Accordingly, students are found to hold a suspicious view toward the effectiveness of content-based AWE feedback, which might influence their subsequent implementation of the feedback. For example, in Li et al.’s (2015) study, some students reported that the AWE feedback at the content level was less useful than the corrective feedback at the linguistic level, and they preferred to seek help from the teachers when it came to problems related to organization, writing strategies, and content creation. Consequently, Li et al. (2015) suggested that additional support in the form of meaningful interaction with humans is essential to support students’ utilization of AWE feedback.

**Collaborative Processing of AWE Feedback**

In response to the above challenges, this study argues that collaboratively processing content-based AWE feedback can be a potentially effective approach to enhance learners’ use of evidence because collaboration may play a positive role in facilitating learners’ knowledge co-construction (Gass, 2003). Studies on collaborative writing have shown that when students write collaboratively, they produce better quality writing, and that the collaborative experience benefits individual writing skill development (Elabdali, 2021; Wigglesworth & Storch, 2009). Thus, collaboration enables co-construction of understanding about writing that can be subsequently transferred to individual writing (Zhang, 2019; see Storch, 2013 for discussion). Following this line of reasoning, we would expect that when students work together to understand what the AWE feedback messages entail and co-decide how to address the feedback and what revisions to make, they might reach a better understanding of the feedback and be able to use it for revision more effectively.

Previous studies have shown that in-class collaborative processing of feedback carries benefit in terms of inducing higher accuracy of error correction (Wigglesworth & Storch, 2012). In one of the few studies that specifically examined collaborative processing of corrective feedback, Kim and Emeliyanova (2019) assigned 36 ESL learners into either the pair-revision condition or individual-revision condition. Participants were provided with indirect written corrective feedback after completing the individual writing tasks. After analyzing the participants’ revision behavior, the researchers found that students who collaboratively responded to the written corrective feedback in pairs corrected their errors at a higher rate of accuracy than those who worked with corrective feedback on their own, possibly because the learners benefited from the occurrence of languaging during pair discussion. Evidence from a recent longitudinal study also suggested that integrating the use of AWE into peer assessment had a positive impact on students’ mindset and motivation to engage with writing (Yao et al., 2021), indicating that peer collaboration might potentially contribute to the utilization of AWE feedback.

Given that collaborative processing involves dialogical discourses around feedback that may induce the co-construction of feedback processing and co-generation of revision strategies (Storch, 2013), it may help
address the de-humanization and lack-of-explicitness issues inherent in content-based AWE feedback. Although previous research has revealed the potential of collaboration in the use of corrective feedback (Elabdali, 2021), these studies primarily documented learners’ collaborative interaction with written feedback in a one-shot design. Little is known about how learners collaboratively make sense of content-based AWE feedback and how this sense-making evolves over time. Driven by the calls to integrate AWE systems with social communication in the classroom, the current study aims to address the following question: How did learners collaboratively process content-based AWE feedback on the use of evidence in argumentative writing over a semester?

Method

Virtual Writing Tutor

In this study, Virtual Writing Tutor (Walker, 2020), a web-based AWE system that can assess the quality of argumentative writing, was used to provide learners with content-based feedback (see Figure 1). Virtual Writing Tutor was chosen because it was one of the few AWE systems that could identify and evaluate the use of evidence in argumentative writing, the target writing element in this study, and is open to public access. Essays submitted to this AWE system will be automatically compared with a built-in argumentative writing outline that includes five paragraphs, namely introduction, first argument, second argument, counter argument and conclusion. Each paragraph will be scored based on the argument-related features.

Figure 1

A Snapshot of Virtual Writing Tutor

When assessing the quality of argument in an essay, the system can detect language related to topic sentence, argue, evidence, cited sources as well as support, and generate a score for each argumentative feature, with highlighted suggestions informing writers how to revise each feature. Table 1 describes the evidence-related features in this AWE system.
Table 1

Evidence-related Features in Virtual Writing Tutor (Walker, 2020)

<table>
<thead>
<tr>
<th>Evidence-related Features</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evidence</td>
<td>Using words/phrases related to providing evidence</td>
<td>“You have not used any words commonly used when giving evidence. That’s not good. Use one or two more words and phrases for giving evidence to get a higher score.”</td>
</tr>
<tr>
<td>Cited Sources</td>
<td>Including an in-text citation</td>
<td>“I was expecting to find a capitalized name (Walker) or year (2019).”</td>
</tr>
<tr>
<td>Support</td>
<td>Using words and phrases for providing support</td>
<td>“Use one or two more support words and phrases for a higher score. Some examples of words that you can use are as follows: a case in point, an analogy…”</td>
</tr>
</tbody>
</table>

Instructional Context

The current study was conducted in an EAP classroom at a university in China. Prior to the enrollment, all students took an English placement test administered by the university. Only those who achieved the intermediate-advanced English proficiency level (roughly equivalent to CEFR B2) in the test were eligible for enrolling in this course. Such a requirement was to ensure that students had the linguistic foundations to use English for academic purposes. Students in this class were of a similar proficiency level. The classroom where we conducted the study consisted of 29 first-year undergraduate students (19 males and 10 females) majoring in electronic engineering.

The purpose of the EAP course was to develop students’ argumentative writing skills. In the beginning of the semester, the teacher taught students the essential elements in composing argumentative essay (claim, evidence, and reasoning) and citation of evidence from credible sources. In each writing assignment, students were first given reading materials to learn how researchers built up evidence-based arguments in research reports. The reading list included the following five topics: (a) the social dilemma of autonomous vehicles, (b) learning math at home, (c) infants’ education, (d) climate change, and (e) transportation policy. After reading these materials, every student was required to construct their own argument in response to the above topics. They were asked to compose an argumentative essay individually on a web-based word processor, Jinshan Document, to state to what extent they agreed or disagreed with the author’s opinion and provide reasons and relevant evidence to support their claims. No requirements were given regarding how many sources to cite as evidence. Each student needed to complete five argumentative writing assignments over the semester.

To strengthen students’ understanding and grasp of evidence use to support claims during argumentative essay writing, we incorporated content-based AWE feedback into the course. To circumvent the various challenges, as reported in literature, that students may encounter when using AWE feedback, we made this step a collaborative activity. In this “collaborative processing of AWE feedback” activity, students were given the autonomy to form groups on a voluntary basis and determine their roles in group collaboration. There were altogether six groups, with four to five students in each group. After each writing assignment, each group was asked to choose one peer’s essay to focus on and work collaboratively to process the AWE feedback. By focusing on a single piece of writing together, each group had a clear focus during the activity and engaged in collective sense-making of the AWE feedback and co-construction of the understanding of evidence use in argumentative essay writing.

The “collaborative processing of AWE feedback” activity consisted of two stages: offline (in-class) group discussion and subsequent online group revision (as shown in Figure 2). During the offline group
discussion, the teacher printed out the content-based AWE feedback on the drafts generated by Virtual Writing Tutor for each group to review and discuss in class. The reason the students were not given direct access to the AWE tool but rather were provided the printed version of the AWE feedback was to ensure that all the discussion around AWE feedback took place in class so that we could fully capture the collaboration process. The participants were first given 5 minutes to read and interpret the feedback individually and then given 10 minutes to share their individual interpretations of the feedback and discuss remedial actions. This timeframe was set based on pilot studies on another cohort of students that were not included in this study and was deemed sufficient for this collaborative task. Students were also required to discuss and collaboratively develop a plan to revise the writing. At this stage, each group was also given a questionnaire to indicate their collective perception toward the usefulness of the AWE feedback based on their collaborative sense-making of the feedback (see Appendix). They were instructed to reach consensus within the group and respond to the evaluation questionnaire as a group.

During the subsequent online group revision, each group had one week to collaboratively revise, asynchronously and synchronously, the draft on Jinshan Document based on the content-based AWE feedback and their revision plan. Each group member had the autonomy to decide their division of labor during collaborative revision. Before submitting the revised draft, all group members were required to proofread the revised writing together and review whether the revised evidence addressed the key issues mentioned in the group discussion. Given that each group had achieved a consensus on the entire revision plan, the division of labor barely influenced the coherence of the argument. The teacher’s involvement was kept minimal during the “collaborative processing of AWE feedback” activity, but the teacher would respond to students’ questions related to argumentative writing, such as the overall structure and rhetoric strategies. The rationale behind this was to avoid interfering with learners’ collaborative sense-making of the AWE feedback so that the participants’ interpretation of the AWE feedback was a collaborative product.

**Figure 2**

*The Two-stage Pedagogical Design*

<table>
<thead>
<tr>
<th>Stage 1</th>
<th>Stage 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group Discussion</strong></td>
<td><strong>Group Revision</strong></td>
</tr>
<tr>
<td>Each group member received and interpreted the AWE feedback individually.</td>
<td>Each group member contributed to the collaborative revision based on the revision plan.</td>
</tr>
<tr>
<td>Each group discussed, evaluated the AWE feedback, and produced a revision plan.</td>
<td>Each group reviewed and proofread the revised draft before submitting it.</td>
</tr>
</tbody>
</table>

**Participants**

The current study adopted a case study approach to investigate learners’ collaborative utilization of content-based AWE feedback to enhance their use of evidence in argumentative writing (Yin, 2018). Purposive sampling strategy was used to recruit participants (Cohen et al., 2013). Since this study aimed at understanding collaborative use of feedback, a group that exhibited a high level of collaboration, operationalized by a high level of equality and mutuality where every group member actively participated in the discussion and share their ideas (Storch, 2002), was selected as the participants. This group was the only group that exhibited the collaborative pattern in the initial three weeks, during which each member equally contributed to the group discussion and responded to their peers’ opinions mutually. Participants’ consent was collected prior to data collection, and pseudonyms were used in the study. Table 2 presents the profiles of the participants. The focal group consisted of five learners (three males, two females), two of
which had experience using a language-based AWE system in high school. As mentioned earlier, these five students had similar English proficiency levels (roughly equivalent to CEFR B2).

### Table 2

Profile of the Participants

<table>
<thead>
<tr>
<th>Name</th>
<th>Gender</th>
<th>Previous experience with AWE systems</th>
<th>English Proficiency (CEFR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jamie</td>
<td>Male</td>
<td>No</td>
<td>B2</td>
</tr>
<tr>
<td>Michael</td>
<td>Male</td>
<td>Yes</td>
<td>B2</td>
</tr>
<tr>
<td>Helen</td>
<td>Female</td>
<td>Yes</td>
<td>B2</td>
</tr>
<tr>
<td>Nick</td>
<td>Male</td>
<td>No</td>
<td>B2</td>
</tr>
<tr>
<td>Laura</td>
<td>Female</td>
<td>No</td>
<td>B2</td>
</tr>
</tbody>
</table>

### Data Collection

In this study, the data were primarily gathered from learners’ online and offline collaborative activities. Initial drafts that received AWE feedback and the corresponding revised drafts in the five writing tasks were retrieved from the web-based writing platform. Their in-class discussions, each lasting 10 minutes, over the five writing tasks were audio-recorded. AWE feedback messages, and their written revision plans were also collected to provide information that would facilitate the interpretation of learners’ collaborative use of content-based AWE feedback. Upon the completion of the group discussion of the revision plan, the group was instructed to discuss and collectively rate the helpfulness of the AWE feedback on evidence, cited sources, and support on a five-point Likert scale (1 = very unhelpful; 5 = very helpful; see Appendix). We collected group evaluation instead of individual evaluation because such response can reflect this group’s collective perception toward the content-based AWE feedback and enable us to interpret students’ collaborative processing of feedback.

Moreover, we also conducted semi-structured interviews in Chinese with each participant to gauge their perceptions toward Virtual Writing Tutor and their learning experiences related to the collaborative use of AWE feedback. In the interview, we asked the participants questions like ‘what was the focus of discussion when you were reading the feedback together’ and ‘how did you revise the evidence collaboratively based on the AWE feedback and why’ to triangulate with the in-class discussion data. The interviews were conducted at the end of the semester to minimize the potential of interfering with the participants’ collaboration behaviors and collaborative learning experience. We were fully aware that leaving the interviews to the end may sacrifice recall quality due to the possibility of lost or distorted memories. To facilitate the participants’ recall of their previous perceptions and experiences, participants’ written texts and the audio-recordings of the discussion were used as the stimuli and constant references to relevant episodes in the discussions and the revision texts were made during the interviews. The interview lasted 20 to 30 minutes for each individual. The overall interview data consisted of 125 minutes of interview responses in total.

### Data Analysis

The participants’ group discussions and interview responses were analyzed to examine their collaborative use of AWE feedback over time (Cohen et al., 2013). First, the discussion and interview data were transcribed verbatim and translated into English. Second, the transcribed data was broken into idea units, an approach that is commonly used in analyzing collaborative discourses (Su et al., 2021). Third, a thematic coding approach was employed to analyze these idea units. With no set coding scheme initially, we coded participants’ discussion and interview data inductively based on their collaborative practices, such as
“comprehending” and “content construction.” Similar codes were later clustered into categories, based on which we determined the themes of collaborative use of AWE feedback. Feedback messages, questionnaires, and revision plans were also analyzed to provide additional information to understand their collaborative decision-making.

We further analyzed and compared the participants’ use of evidence in the first and revised drafts to examine whether the collaborative engagement with content-based AWE helped enhance the use of evidence in argumentative writing. First, we manually coded the evidence in each draft based on McNeill and Krajcik’s (2009) conceptualization, who defined the use of evidence as the information used to justify a claim in the argument. The citation of textual evidence from the reading materials was not counted as instances of evidence use because it only involved the summarization of the authors’ opinions instead of the selective use of evidence to support their own claims. We further coded individual participants’ contribution to the construction of the evidence. In total, we coded 18 pieces of evidence in the five writing assignments.

Second, two researchers assigned codes independently to the extracted evidence by adapting Jin et al.’s (2019) framework of evidence use, in which the effectiveness of evidence use is assessed through incorporation, elaboration and reasoning (see Table 3). The sources of evidence were also manually coded to evaluate their credibility. For example, if a piece of evidence was cited from a research report and was clearly verifiable, this piece of evidence would be coded as “credible,” whereas a piece of evidence that vaguely described a social phenomenon that could not be verified through sources would be coded as “not credible.” The inter-coder agreement of the coding was over 92%, and the discrepancies were discussed and solved. Students’ use of evidence and the sources of evidence were then compared across the five writing tasks.

Table 3

Coding Scheme of Evidence Use (Adapted from Jin et al., 2019)

<table>
<thead>
<tr>
<th>Evidentiary Practice</th>
<th>Feature</th>
<th>Description</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incorporation</td>
<td>Relevance</td>
<td>The evidence is relevant to the claim.</td>
<td>Relevant, Irrelevant</td>
</tr>
<tr>
<td></td>
<td>Credibility</td>
<td>The evidence comes from credible sources that can be verified.</td>
<td>Credible, Not credible</td>
</tr>
<tr>
<td>Elaboration</td>
<td>Detailedness</td>
<td>The evidence is elaborated on in detail.</td>
<td>With details, No details</td>
</tr>
<tr>
<td>Reasoning</td>
<td>Analysis</td>
<td>A clear link between evidence and claim is established.</td>
<td>With analysis, No analysis</td>
</tr>
</tbody>
</table>

Results

It was found that the participants’ collaborative use of content-based AWE feedback on evidence use went through three stages, which corresponded with their changing perceptions of the usefulness of AWE feedback. As reflected in Figure 3, participants’ perceptions of the helpfulness of AWE feedback fluctuated among Unhelpful, Neutral, and Helpful over the five writing assignments.
Figure 3

Perception of the Helpfulness of Different Aspects of AWE Feedback

Note. The perception of helpfulness was rated based on the group’s consensus.

In the initial phase (Assignment 1), participants were fond of the automated feedback and perceived the feedback on evidence and support as helpful. This phase is coded as the “trustful” phase. After two weeks of exposure to the AWE feedback, the participants entered the “skeptical” phase (Assignment 2, 3, and 4), where they expressed a level of uncertainty about the usefulness of AWE feedback on evidence, cited sources, and support. In the final stage, the “critical” phase (Assignment 5), the participants were more critical of the automated feedback and expressed that AWE feedback on evidence, cited sources, and support were all unhelpful (see Table 4).

Table 4

Number of Pieces of Evidence

<table>
<thead>
<tr>
<th>Assignment</th>
<th>First Draft Author</th>
<th>Evidence in First Draft</th>
<th>Evidence in Revised Draft</th>
<th>Evidence-related Revision</th>
<th>Participants who contributed to the discussion on the revision of evidence</th>
<th>Participants who made evidence-related revisions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Michael</td>
<td>1</td>
<td>4</td>
<td>Added 4 sentences</td>
<td>Helen, Laura</td>
<td>Helen, Laura, Michael</td>
</tr>
<tr>
<td>2</td>
<td>Laura</td>
<td>1</td>
<td>1</td>
<td>None</td>
<td>Jamie, Laura, Michael</td>
<td>None</td>
</tr>
<tr>
<td>3</td>
<td>Helen</td>
<td>1</td>
<td>1</td>
<td>Added 2 sentences; Deleted 2 sentences</td>
<td>Helen, Jamie, Laura</td>
<td>Helen, Michael</td>
</tr>
<tr>
<td>4</td>
<td>Nick</td>
<td>1</td>
<td>2</td>
<td>Added 1 sentence</td>
<td>Jamie, Laura</td>
<td>Helen, Nick</td>
</tr>
<tr>
<td>5</td>
<td>Michael</td>
<td>2</td>
<td>4</td>
<td>Added 5 sentences</td>
<td>Jamie, Michael</td>
<td>Helen, Nick</td>
</tr>
</tbody>
</table>

Note. There were no evidence-related revisions in Assignment 2 because the system failed to identify the evidence and provided no feedback.
Trustful Phase: Collaborative Construction of the Structural Elements of Evidence Use

In the trustful phase, the participants’ evaluation of the helpfulness of the AWE feedback was in general positive, rating feedback on evidence and support both useful (4 out of 5). Their interview responses around the feedback also showed a trustful attitude. For instance, Jamie recounted, “The automated feedback was helping a lot in the beginning, because the system accurately identified my weaknesses and showed me what an argumentative essay was like.” There was no evidence of the questioning of the feedback in the participants’ collaborative discourses. Rather, their collaborative discourses centered around understanding the terminologies appeared in the AWE feedback messages and comprehending the difference between these evidence-related elements (see Excerpt 1). The participants showed curiosity in some of the terminologies and noticed the three evidence-related features given by Virtual Writing Tutor, but seemed to be uncertain of the precise definition of evidence, cited sources, and support.

**Excerpt 1** (Assignment 1)

Laura: Evidence, cited sources and support, these are all argumentative elements. What is the difference between evidence and support though?
Helen: I think evidence includes cited source and support.
Jamie: Cited sources refers to quotation, indicating the source.
Laura: I agree.
Michael: Does evidence equal to data?
Nick: Perhaps evidence is the process of reasoning.

The participants’ collaborative discourses primarily focused on the structure of evidence use. This focus had something to do with the participants’ limited understanding of the argumentation structure. According to Nick,

> It was my first time writing an argumentative essay in English, and I knew little about its writing structure. I didn’t know there were a number of argumentative elements, such as argue, evidence, reasoning, that should be included in my writing until I received the AWE feedback.

The AWE feedback for the group’s Assignment 1 suggested that they did well in providing a topic sentence and a clear claim, but did a poor job in providing evidence, citing sources, and supporting the argument (see Figure 4). Indeed, the lack of evidence in the first draft of Assignment 1 revealed that they did not have a comprehensive understanding of the use of evidence. The automated feedback, therefore, played a role in drawing the participants’ attention to the importance of evidence use, highlighting several evidence-related features that the participants might not be aware of previously. This phenomenon was also confirmed by Helen, who reported, “The automated feedback was helpful in the beginning, because it let me know that using evidence should include credible sources and a reasoning process.”
Getting access to information previously unknown made the participants take the feedback uncritically, and their collaborative discourse appeared to be largely homogeneous, without much dispute, in this trustful phase. Instead, they focused primarily on working together to make sense of the terminologies in the AWE feedback. Given their limited prior knowledge of argumentation and the use of evidence, the participants found the feedback message related to the use of evidence perplexing. Michael observed that, “I was unsure about the precise definition of evidence in the AWE feedback, and I would be unable to revise my use of evidence without understanding what the AWE system wanted me to do.” Thus, the lack of relevant understanding made the AWE feedback inaccessible and unusable to some participants. Accordingly, the participants’ collaboration primarily focused on co-constructing understanding of structural elements of evidence use featured in the feedback and brainstorming how they could respond to the feedback.

Their revisions mainly focused on the structure of evidence as well. The AWE feedback made them realize that their initial evidence use was incomplete, and it was important to include relevant and credible evidence in the argument. As shown in the revision plan that they co-constructed after the discussion (Figure 5), the participants noted down including more data and evidence, which revealed the consolidation of their collaborative conceptualization of evidence use.
The revision records online also showed that participants focused their revision on expanding the sources of the evidence, presenting multiple pieces of evidence to enrich their argument (see Figure 6). Although each participant revised a different section of the essay, their revision behaviors were quite homogenous, primarily focusing on the structural elements of the evidence, and centered around incorporating multiple pieces of evidence to make the argument complete. For instance, in the revised draft, the participants added people’s conservative attitude towards Autonomous Vehicles (AVs) and transgenic soybeans as well as the “Moral Machine” experiment as evidence to substantiate the claims in their initial draft. Thus, both their group discussion and revision indicated that the participants’ primary focus was on constructing a complete argument, presenting evidence right after putting forth a claim.

**Figure 6**

*Evidence Comparison in Assignment 1*

[Skeptical Phase: Collaborative Attention to the Content of Evidence Use]

Starting from Assignment 2, the participants’ collaborative discourse around AWE feedback shifted from a primary attention to structural elements to an increasing focus on the content of evidence use. Instead of discussing the key terminologies related to the structure of evidence use, the participants paid more attention to the content of the evidence use, such as clarifying the relevance between evidence and claim and enriching the details. However, they also pointed out the fact that the AWE system did not accurately identify their evidence use (see Excerpt 2).

**Excerpt 2** (Assignment 2)

Laura: That’s weird. The system didn’t identify the evidence I used here. I even cited sources. But it says I didn’t.

Jamie: Right, I think your evidence is well-supported though.

When discussing the AWE feedback for Assignment 3, the participants pointed out the need to emphasize how Clinton’s experience was related to learning to give up. This indicates that the participants’ primary focus of developing the use of evidence transitioned to constructing content in each evidence-related element, namely incorporation, elaboration, and reasoning. As shown in Excerpt 3, Laura drew peers’ attention to specific feedback related to the content of evidence use (i.e., “had not used any word to support your claim”), which led to Helen’s critical reflection on reasons behind the feedback, and related
collaborative discourses around clarifying the claim and devising ways to make the relationship between the evidence and the claim more salient.

**Excerpt 3** (Assignment 3)

Laura: I agree that the reasoning process is absent. The system suggests that you had not used any word to support your claim.

Helen: That is true. I used Clinton’s experience to enrich my argument, but I didn’t explain the connection between the evidence and my claim.

Jamie: What is your claim exactly?

Helen: I argue that it is also important to learn to give up sometimes.

Laura: Then we need to revise the content and make it more concise and relevant to the claim.

Thus, the participants built on each other’s understanding and analysis, and successfully constructed a consensual direction for revision that focused on enhancing the relationship between the evidence and the claim. The shifted collaborative focus towards the content of evidence use was also reflected in the participants’ revision behaviors. Figure 7 summarized the revisions the participants made during Assignment 3. It showed that their revisions were primarily about deleting the details that they deemed as irrelevant to the claim and linking evidence with the claim to enhance the relevancy. However, despite their efforts in enhancing the content quality of the evidence use, some of their revisions seemed to be ineffective. For instance, the participants did not provide the context of Clinton’s speech and indicate why giving up was a right choice for Clinton’s career path. Most importantly, the claim was related to infants’ education. But the participants did not address the similarity between adults’ life choices and infants’ education.

**Figure 7**

*Evidence Comparison in Assignment 3*

The participants attributed the ineffective revision to the limitation of the feedback provided in the AWE system. For instance, Helen recounted in the interview,

We began to question the usefulness of the automated feedback because it was rigid. The feedback messages appeared to be similar all the time, such as ‘you had used words related to evidence, that is good’. In fact, we were more interested in learning how to improve the content of our evidence. But the system couldn’t provide such feedback.

The limited feedback provided by *Virtual Writing Tutor* explained why the participants failed to make effective revision to improve content quality, even though they attempted to do so.

The participants’ shifted attentional focus from the structural elements to the content of evidence use, coupled with the lack of explicitness and concreteness in the feedback on this aspect provided in the AWE system, induced the participants’ skeptical mentality towards the AWE system. Given their familiarization with the structural elements of evidence use, they found AWE feedback that indicated whether they had used evidence or not was no longer helpful. Since they had not grasped the elaboration and reasoning
dimensions of evidence use, the participants still found it helpful when the AWE system informed them that a reasoning process was missing. However, they found that the feedback message was inadequate in pinpointing the exact issues and providing detailed diagnosis concerning these dimensions, and it did not give constructive and concrete suggestions to support the revision on these dimensions. Moreover, the participants found the recommended linguistic expressions related to the use of evidence provided in the Virtual Writing Tutor helpful. Even though such feedback was repetitive and concentrated solely on the linguistic forms, Michael mentioned that “Using the recommended expressions provided the opportunity to elaborate on my writing. I used to write very little about evidence, but when I was using those phrases, I had to write more details in order to complete the sentence.” In Assignment 4, the participants used the phrase “according to” twice to elicit more details (see Figure 8). For example, in the original evidence, the author provided concrete information regarding the report, which included the publisher, the date, and comments on the report. The added evidence in the revised draft also included details such as the exact number of extinct species. Such specific information indicated that the participants performed well in elaboration through using the recommended linguistic expressions.

**Figure 8**

Evidence Comparison in Assignment 4

The participants’ skeptical mentality towards the feedback they received in the AWE system shaped their heterogeneous focus during in-class discussion and while collaborating asynchronously online to revise the draft. As Laura said, “Everyone’s knowledge about using evidence is different. Some are good at providing evidence sources, while others perform well in reasoning.” Individual learners tended to focus on different aspects of using evidence during collaboration primarily because the focus on content construction covers multiple aspects of the use of evidence, which included justifying the relevance and credibility in incorporation, providing details in elaboration, and constructing an analysis process in reasoning. Each participant’s understanding of and ability to grasp these aspects varied, and collaborative use of AWE feedback enabled the participants to contribute individual strengths to the revision process. This collaboration process based on distributed expertise led to potential learning opportunities for individual participants.

**Critical Phase: Collaborative Consideration of the Persuasiveness of Evidence Use**

As the participants moved to the critical phase, their collaborative discourse no longer focused on the AWE feedback, as they deemed the automated feedback inaccurate. Indeed, in the first draft of Assignment 5, the writer presented a long piece of well-elaborated evidence, with relevant and credible source, sufficient details and some analysis. But Virtual Writing Tutor failed to identify the evidence and suggested that evidence, cited sources, and support were missing. Therefore, in the discussion of Assignment 5, the participants focused more on the overall persuasiveness of using evidence, an aspect that the AWE system did not address but was perceived crucial by the participants. As shown in Excerpt 4, Jamie and Nick were discussing a possible way to strengthen the persuasiveness of the evidence.
Excerpt 4 (Assignment 5)

Jamie: You used only one example to describe the policy in a developing country. That is not enough.
Nick: What is your suggestion?
Jamie: You can add a contrastive example in a developed country to illustrate the difference.
Nick: That could be a good idea. I can describe the transportation policy in Tokyo.
Michael: That’s great! You can also focus on explaining the relationship between population and transportation policy.
Nick: The argument will sound more logical then.

Even though the initial evidence in Assignment 5 was relevant to the claim, quoted from a credible source with concrete details and reasoning, Jamie suggested that there was still room for improvement to enhance the overall persuasiveness of the evidence. Given that Virtual Writing Tutor did not require learners to incorporate more evidence, the suggestion of adding a contrastive example could be seen as the learners’ original consideration out of their collaborative use of AWE feedback. As shown in Figure 9, the participants had presented a vivid illustration about the reasoning process. In an attempt to enhance the overall persuasiveness, the participants planned to clarify the relationship between population in Jakarta, traffic problems, and the size of the city.

Figure 9

Revision Plan for Assignment 5

The collaborative consideration of persuasiveness at this phase is indicative of learners’ development of using evidence. Their basic understanding of evidence had been reinforced. Participants were able to present a piece of complete evidence in the draft. They also became capable of providing concrete details and relatively sound reasoning in the evidence. Jamie observed that, “In the beginning, I was unfamiliar with the concept of using evidence, so I had to follow the system’s instruction to use and revise evidence, such as adding examples. Now that I became familiar with the structure of using evidence and aware of the importance of including relevant details, I consider more about the readability, that is how to better convince the readers.” The participants’ focus on the persuasiveness of the evidence use was also reflected in their revision behaviors. Figure 10 presents the revised evidence in Assignment 5.
The highlighted sentences on the right side were additional pieces of evidence used to create a contrastive effect to strengthen the persuasiveness of the essay. The authors first cited an additional report to illustrate the large population in Jakarta and pointed out that the previous public transportation was unable to cope with the population. Then, the authors mentioned London’s subway to make a comparison with Jakarta’s railway. Albeit short in length, the example of the subway network in London included a precise number to illustrate how well-managed the transportation system was. Moreover, this piece of evidence was inserted into the evidence of Jakarta’s policy instead of being presented separately in the following paragraphs. Even though the evidence about London’s subway was not well connected with the following pronoun reference, such arrangement confirms that the participants were engaging with higher-level thinking and transferring their collaborative consideration into writing. At this stage, the participants found the overall AWE feedback on the use of evidence ineffective, since providing informative feedback on the persuasiveness of evidence was beyond the capability of the AWE system.

In general, the results indicate that effective collaboration played a positive role in facilitating students’ utilization of the AWE feedback (see Table 5). Initially, the students primarily attended to the structure of using evidence as well as its elements. After developing the basic understanding of using evidence, the students’ focus on evidence gradually shifted from the structure to the content and the overall persuasiveness.
## Table 5

**Comparison of Evidence Use across Five Assignments**

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Draft</th>
<th>Evidence Source</th>
<th>Incorporation</th>
<th>Elaboration</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>First</td>
<td>Social Phenomenon</td>
<td>Relevant but Not Credible</td>
<td>No details</td>
<td>No Analysis</td>
</tr>
<tr>
<td></td>
<td>Revised</td>
<td>Social Phenomenon</td>
<td>Relevant but Not Credible</td>
<td>No details</td>
<td>No Analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>News*</td>
<td>Relevant and Credible</td>
<td>With details</td>
<td>No Analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Research Report*</td>
<td>Relevant and Credible</td>
<td>With details</td>
<td>No Analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Research Report*</td>
<td>Relevant and Credible</td>
<td>With details</td>
<td>No Analysis</td>
</tr>
<tr>
<td>2</td>
<td>First</td>
<td>Research Report</td>
<td>Relevant and Credible</td>
<td>With details</td>
<td>No Analysis</td>
</tr>
<tr>
<td></td>
<td>Revised</td>
<td>Research Report</td>
<td>Relevant and Credible</td>
<td>With details</td>
<td>No Analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Quotation</td>
<td>Relevant but Not Credible</td>
<td>With details</td>
<td>No Analysis</td>
</tr>
<tr>
<td></td>
<td>Revised</td>
<td>Quotation</td>
<td>Relevant but Not Credible</td>
<td>No details</td>
<td>No Analysis</td>
</tr>
<tr>
<td>3</td>
<td>First</td>
<td>Research Report</td>
<td>Relevant and Credible</td>
<td>With details</td>
<td>With Analysis</td>
</tr>
<tr>
<td></td>
<td>Revised</td>
<td>Research Report</td>
<td>Relevant and Credible</td>
<td>With details</td>
<td>With Analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Research Report*</td>
<td>Relevant and Credible</td>
<td>With details</td>
<td>No Analysis</td>
</tr>
<tr>
<td>4</td>
<td>Initial</td>
<td>Social Phenomenon</td>
<td>Relevant and Credible</td>
<td>With details</td>
<td>With Analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>News</td>
<td>Relevant and Credible</td>
<td>No details</td>
<td>No Analysis</td>
</tr>
<tr>
<td></td>
<td>Revised</td>
<td>Social Phenomenon</td>
<td>Relevant and Credible</td>
<td>With details</td>
<td>With Analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>News</td>
<td>Relevant and Credible</td>
<td>No details</td>
<td>With Analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Research Report*</td>
<td>Relevant and Credible</td>
<td>With details</td>
<td>No Analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Social Phenomenon*</td>
<td>Relevant and Credible</td>
<td>With details</td>
<td>No Analysis</td>
</tr>
</tbody>
</table>

*Note.* Evidence retrieved from additional sources during collaborative revision.

## Discussion

This case study examined a group of five undergraduate learners’ collaborative use of content-based AWE feedback on their use of evidence in argumentative writing. It found that the participants’ use of evidence did improve over time. In contrast with Wang et al.’s (2020) study where the majority of students made ineffective revisions based on the AWE feedback, the five participants in this study demonstrated improvement in their use of evidence. The study found that collaborative discourse around AWE feedback contributed to the observed enhancement of evidence use. When learners were given the opportunity to exchange ideas and co-construct content when revising their use of evidence, they expanded their understanding in terms of incorporating evidence from various sources, enriching the details, building up a reasoning process, and considering the overall persuasiveness. Thus, consistent with Kim and Emeliyanova’s (2019) study, the collaborative use of feedback can help enhance the efficacy of feedback for writing development. The findings suggest that collaborative use of AWE feedback might be a viable pedagogical scaffolding that teachers may utilize to enhance the use of AWE feedback for writing development.
This study took a longitudinal perspective to examine the participants’ collaborative use over time. The study found that the learners’ perception and use of the AWE feedback changed over time, where they shifted from a trustful and dependent relationship with the feedback in guiding the revision to a more critical mentality towards the value of the feedback. In the critical phase, learners were discovered to exhibit greater reliance on collective efforts in interpreting and devising revision solutions to address the areas of weaknesses identified by the system. Consistent with Ranalli’s (2021) findings in which trust-related issues played a critical role in students’ engagement with AWE feedback, this study reveals that learners’ perception toward the AWE system was associated with their collaborative focus on the use of automated feedback. The change of perception was induced by their developing knowledge and skills related to evidence use in argumentative writing and the (mis)match of the affordances of AWE feedback and their developing capacities and changing needs for feedback over time. It seems that collaboration played an increasing role in realizing the potentials of AWE feedback for writing development, as learners progressed to the more sophisticated stage of evidence use in argumentative writing. This longitudinal view of learners’ collaborative use of AWE feedback suggests adopting a dynamic approach towards the pedagogical use of AWE feedback and the types of support that need to be built in at different stages to maximize the potential of AWE feedback.

The study identified three progression phases of learners’ collaborative evidence use, from a focus on structural elements of evidence use to an attention to the quality of evidence use. Students need different support at different stages and the changing support needs call for a coordination of human and AWE scaffolding. At the developing knowledge about structural elements of evidence use stage, some learners might find some terminologies contained in the AWE feedback perplexing without direct explanation, and others may develop inaccurate understanding of these content-related features. Thus, teachers may provide explicit explanations of the terminologies that may appear in AWE feedback to better student comprehension. As learners progress to the stages where they can attend to the quality of evidence use, the feedback provided by AWE might be limited in facilitating the enhancement of detailedness and overall persuasiveness needed. Collaborative discourses are essential at such stages for learners to brainstorm revision strategies. Teachers may need to pay special attention to structure the collaborative tasks in ways that would facilitate effective collaboration at those stages. Moreover, as it has been suggested that teachers’ explicit instruction can be helpful in solving learners’ problems in collaborative argumentative writing (Jin et al., 2020), teachers’ complementary suggestions can optimize the implementation of AWE system in writing instruction.

Conclusion

This study shed light on how a group of five intermediate-advanced EFL learners collaborated around AWE feedback in strengthening the evidence use in their L2 writing over time. It found that collaborative processing of content-based AWE feedback helped strengthen the participants’ understanding of and ability to utilize AWE feedback for writing revision, and that collaborative discourse was more essential to successful revision as the revisions progressed towards the more sophisticated aspects of evidence use over time. Specifically, content-based AWE feedback might be effective in supporting learners’ development of the basic knowledge about evidence use (e.g., the structure of a complete argument as well as the elements in using evidence, such as cited sources and reasoning), but collaboration was observed to be essential to the development of more sophisticated understanding of evidence use (e.g., the relevance and persuasiveness of evidence use). Thus, learners’ progressive collaborative patterns were associated with their developing understanding and skills of evidence use, as well as feedback needs at different developmental stages and what the AWE feedback could afford. The findings of the current study suggest that collaborative use of AWE feedback might be a useful and necessary pedagogical practice to enhance the educational potential of content-based AWE feedback for augmentative writing instruction, especially with regard to more sophisticated aspects of augmentative features.
Owing to the nature of all case studies, this study also has some limitations. First, we only presented one group of learners’ collaborative use of AWE feedback. Their learning experiences and collaborative patterns might have biased our findings on the progressive development of knowledge and skills related to evidence use and the effects of collaboration on their knowledge and skill development. Although the findings indicated what collaborative sense-making of content-based AWE feedback could possibly achieve, they might not be generalizable to other groups that exhibit less collaborative patterns (e.g., dominant/dominant, expert/novice). A larger scale study that involves quantitative analysis will yield more robust and generalizable research findings. Moreover, we only captured the collaboration among university learners who were of intermediate-advanced English proficiency level and had some prior experience with AWE system. The collaborative utilization of content-based AWE feedback may be different among learners who have no prior experience with AWE, who are at a lower proficiency level, or who are at different levels of cognitive maturity (e.g., secondary school students), since learners’ sense making of AWE feedback may be subject to their language proficiency and cognitive abilities. It should also be acknowledged that the content-based AWE system implemented in this study only provides three evidence-related features and generic feedback messages. AWE systems that incorporate other evidence-related features and can generate more specific feedback messages may have different impacts on learners’ collaborative processing of feedback and development of evidence use. Furthermore, there were six groups of learners in this class, but only one group (less than 20%) exhibited collaborative use of AWE feedback despite the opportunities for collaboration built in at each stage of writing. It would be insightful to conduct multiple case studies (e.g., Koltovskaia, 2020; Ranalli, 2021) to examine the factors that facilitate or constrain collaborative AWE feedback use, and to contrast the revision behaviors of groups that exhibited different levels of collaboration when engaging with AWE feedback.

Future research may consider extending this work to investigate the impact of content-based AWE feedback on learners’ use of evidence over different phases. In addition, cross-sectional studies can also be done to examine how patterns and quality of collaborative discourses around AWE feedback might affect the impact of collaboration both on the collaborative work itself and individual learners’ development of relevant knowledge and skills. Researchers can also compare the effect of AWE-only and AWE-combined-with-human-feedback pedagogical approach on learners’ development of evidence use, which can enrich our understanding of how to optimize the potential of content-based AWE feedback.

Acknowledgements

This research was supported by a grant from the National Social Science Fund of China (18BYY110) to the corresponding author of this article (Tan Jin). We would like to express our gratitude to the editors and the anonymous reviewers for their constructive comments and suggestions. Particular thanks also go to the students who participated in this study.

References


Appendix. Questionnaire

Response to Feedback

Group: ___________  Leader: ___________

Please rate the helpfulness of the AWE feedback on the basis of the following:

Claim
1. How would you rate the AWE feedback on Thesis in the introduction?
   Very Unhelpful  Unhelpful  Neither Helpful Nor Unhelpful  Helpful  Very Helpful

2. How would you rate the AWE feedback on Topic Sentence in the arguments?
   Very Unhelpful  Unhelpful  Neither Helpful Nor Unhelpful  Helpful  Very Helpful

3. How would you rate the AWE feedback on Argue in the arguments?
   Very Unhelpful  Unhelpful  Neither Helpful Nor Unhelpful  Helpful  Very Helpful

4. How would you rate the AWE feedback on Reformulation of the Thesis in the conclusion?
   Very Unhelpful  Unhelpful  Neither Helpful Nor Unhelpful  Helpful  Very Helpful

Evidence
1. How would you rate the AWE feedback on Evidence in the arguments?
   Very Unhelpful  Unhelpful  Neither Helpful Nor Unhelpful  Helpful  Very Helpful

2. How would you rate the AWE feedback on Cite Sources in the arguments?
   Very Unhelpful  Unhelpful  Neither Helpful Nor Unhelpful  Helpful  Very Helpful

Reasoning
1. How would you rate the AWE feedback on Support in the arguments?
   Very Unhelpful  Unhelpful  Neither Helpful Nor Unhelpful  Helpful  Very Helpful
About the Authors

Zhan Shi holds an MPhil degree in Education from University of Cambridge. He is pursuing his PhD in Education at The University of Hong Kong in September 2022. His research interest primarily focuses on technology-assisted language learning.

E-mail: shi.zh@outlook.com

Fengkai Liu is a graduate student pursuing his PhD in the department of Linguistics and Translation at City University of Hong Kong. His research interest primarily focuses on computational linguistics.

E-mail: fengkaliu3-c@my.cityu.edu.hk

Chun Lai is an Associate Professor at the Faculty of Education, The University of Hong Kong. Her research interest is technology-enhanced language learning. Her recent research focuses on self-directed language learning with technology beyond the classroom.

E-mail: laichun@hku.hk

Tan Jin (corresponding author) is a Professor in the School of Foreign Studies at South China Normal University. His research interests include corpus linguistics, language testing and assessment, as well as computer assisted language learning.

E-mail: tjin@scnu.edu.cn