Automated Writing Evaluation: Students’ Perceptions and Emotional Involvement

Mei-jung Wang
National Kaohsiung University of Hospitality and Tourism
sebrina@mail.nkuht.edu.tw

David Goodman
National Kaohsiung University of Hospitality and Tourism
davidg@mail.nkuht.edu.tw

Abstract
Breakthroughs in educational technology have made the individualized instruction and immediate feedback features of automated writing evaluation (AWE) increasingly attractive to language teachers. Many previous studies on AWE have focused on psychometric evaluations of its validity, while others have dealt with pedagogical practices and student perceptions of learning effectiveness. Affective considerations—particularly the potential of writing apprehension resulting from the use of AWE—have not been well investigated. This study aims to explore EFL (English as Foreign Language) students’ perceptions of the feedback they receive from AWE, as well as the emotions that accompany the AWE process. The participants included forty-six students who took a writing course. The study compared and examined students’ responses to teacher feedback, peer feedback, and AWE feedback. In addition, the emotions involved in the process of using AWE were examined. The results provide teachers with additional perspectives regarding the use of AWE in EFL writing classes.

Key Words: writing feedback, automated writing evaluation, emotions
INTRODUCTION

Breakthroughs in educational technology have made the individualized instruction and immediate feedback features of automated writing evaluation (AWE) increasingly attractive to language teachers. H. J. Chen, Chiu, and Liao (2009) indicated that feedback plays a central role in second language (L2) writing pedagogy because students can incorporate feedback into the revision of their writing, provided the feedback is specific and clear. Teaching second language writing in Taiwan, however, offers a number of challenges with regard to the provision of feedback to learners, especially in large class settings. The workload involved in correcting and grading students’ writing is substantial, presenting a further pedagogical hurdle in addition to the fact that English as Foreign Language (EFL) students often find English composition classes to be a source of anxiety.

Most studies on AWE have focused on psychometric evaluations of its validity, though some have dealt with pedagogical practices and student perceptions of learning effectiveness. The affective aspects of the use of AWE, including such emotions as writing anxiety, as well as positive emotions or technological-related apprehensive or negative feelings towards writing, have not been well investigated. In 2011 the AWE software system ETS Criterion was introduced and integrated into the writing classes at our university, and thereby supplemented the existing use of teacher-based and peer-based feedback. This study aims to explore student perceptions of the different types of feedback, as well as the emotions that accompany the AWE process.
LITERATURE REVIEW

Different types of writing-based feedback can be provided to students in English as a Foreign Language (EFL) settings. Which types of feedback to make use of and in what manner is feedback to be delivered are matters which continue to attract a great deal of attention in the area of EFL writing research. In order to provide appropriate feedback, teachers must consider factors such as the objectives of the writing class, the background and level of the students, the teaching circumstances and the available resources. In addition to instructor and peer feedback, improvements in information technology have made computer assisted instruction increasingly feasible, and as a result both teachers and students have increasingly shown interest in exploring feedback options which make use of this technology. Some studies (e.g., Fang, 2010; Hay & Isbell, 2000) have even shown that students prefer writing activities with computer-assisted writing software, as long as it is not overly complicated.

Recently, a number of researchers (C.-F. Chen & Cheng, 2006; H. J. Chen, 2006; H. J. Chen et al., 2009; Otoshi, 2005; N. D. Yang, 2004) have focused on the validity of software-based scoring of writing (e.g., H. J. Chen, 2006; H. J. Chen et al., 2009). H. J. Chen et al. (2009) examined the grammar feedback provided by two well-known systems: Vantage My Access and ETS Criterion. Both systems provide a wide range of resources and features, including model essays and progress reports. The study showed, however, that neither system provided satisfactory feedback to EFL students. Although the My Access program provides approximately 30 different types of feedback message, many of these are false error
messages. The Criterion program performed better, but it still failed to address common errors related to word order, modals, tenses, collocations, conjunctions, word choice, and pronouns.

Still other studies have investigated student perceptions of the integration of AWE into the writing classroom. Fang (2010), for example, investigated the perceptions which EFL learners held toward AWE in the context of a college composition class. Survey results showed that the majority of the learners held positive views regarding the use of AWE as a writing instruction tool, particularly for feedback related to form. Conversely, learners were less positive concerning the use of AWE as an essay grading tool. Another study, conducted by Lai (2010), compared learner perceptions of the effectiveness of AWE with that of peer evaluation (PE). Twenty-two EFL English majors received feedback from AWE and were asked to make revisions based on that feedback. In addition, they also made revisions on the basis of peer feedback. Survey results showed that the students considered feedback from AWE to be less helpful than PE, and so they opted for PE over AWE. C.-F. Chen and Cheng (2006, 2008) also explored factors underlying the perceived effectiveness, or lack thereof, of AWE in the context of three teachers using an AWE program in their EFL writing classes. They found that student perceptions regarding the usefulness of AWE were influenced by attitudes toward whether or not the automated scores were seen as fair, whether the diagnostic feedback was held to be informative, and the manner in which the teachers integrated AWE into the writing class.

The results above suggest that feedback from AWE can assist teachers in the time-consuming work of writing instruction, but that its use cannot supplant the role of the teacher, and indeed of peers, in
providing useful feedback. Most researchers advise that AWE software should not be used in the classroom as a stand-alone evaluation tool (Lai, 2010; Otoshi, 2005), and furthermore recommend a thorough analysis of the pedagogical needs and aims of the particular context before using AWE in order to avoid inappropriate use or undesirable consequences (C.-F. Chen & Cheng, 2006, 2008). As Guénette (2007) has pinpointed, teachers must not forget that learning a second language requires time, effort, and patience, and that corrective feedback is only one of the many factors involved in the process.

Among potential unintended consequences of adopting an AWE program, writing apprehension and negative attitudes arising from the use of educational technology are particularly worthy of attention. In recent decades, the phenomenon of writing anxiety has attracted attention from researchers, both in first language (L1) and L2 writing, but particularly in the latter. Students who experience writing anxiety tend to feel an inability to organize their ideas coherently (Schweiker-Marra & Marra, 2000). Low self-confidence and low English proficiency seem to play a major role in creating anxiety—a key reason why so much attention is paid to L2 writing in this regard—with these deficiencies negatively influencing learners’ achievement (Bailey, Daley, & Onwuegbuzie, 1999; Cheng, 2004; Erkan & Saban, 2011; Tsai & Cheng, 2009).

Research related to negative attitudes arising from the use of educational technology (e.g., Järvenoja & Järvelä, 2005; Nummenmaa, 2007; Wosnitza & Volet, 2005) shows that affective reactions to technology-enhanced learning environments are influenced by self, others, context, task, and the nature of the
technology itself. Wosnitza and Volet (2005), for example, pointed out that individuals’ inability to handle technological challenges can lead to negative emotional reactions, as can technological failure or peers’ lack of enthusiasm or commitment to an online exchange. Nummenmaa (2007) emphasized that emotions are important determinants of student behavior in web-based learning, while at the same time pointing out that the technology itself is not the only determinant of students’ emotional reactions.

On the positive side, Dunkel (1990) claimed that computer technology as a pedagogical tool could potentially increase language learners’ (1) self esteem, (2) vocational preparedness, (3) language proficiency, and (4) overall academic skills. The hope is that incorporating information communication technology (ICT) into language instruction can be successful if undertaken with the use of the correct strategies, but instructors and learners need to be aware of the potentially negative emotional reactions. Overall, relatively few empirical studies have examined the extent to which different ICTs positively or negatively affect learners’ emotions (C. M. Chen & Wang, 2011) and there is still a relative dearth of studies focusing on the emotions which teachers and students might experience when using ICT in the classroom (Kay & Loverock, 2008).

The studies mentioned above provide evidence regarding students’ responses to AWE programs in language learning classrooms and highlight the possible negative consequences of incorporating AWE programs. However, previous studies have mostly focused on psychometric evaluations of the validity of automated writing evaluation (AWE) and comparisons of AWE and human writing evaluation. The current study extends previous
research by addressing students’ views of different kinds of writing feedback, and taking a more nuanced look at the range of emotions which students may encounter when using AWE. In response to these problems, this study explores the following research questions:

1. What are students’ perceptions of teacher, peer, and AWE feedback, particularly the degree to which learners feel that these types of feedback facilitate improvements in their writing?

2. What emotions do students experience as they are exposed to AWE in their writing class?

METHODS

Subjects

Forty-six students from the Department of Applied English in a university in Taiwan participated in this study. They were second-semester university freshmen at the time of the study, and the writing course was a requirement within the departmental curriculum. All first-year and second-year writing courses in the department are conducted in groups of 30 or fewer students in order to allow the instructor to provide more feedback to the students.

For this study, the students were taught in two groups by two different instructors. Although an experimental design was not strictly applied, it bears pointing out that the division into groups was not based upon proficiency level or other non-arbitrary student characteristics, and that the teachers attempted to use the same methods of instruction. The data elicited for use in the study did not require the instructors to change the curricular design of the course,
and so the research design had minimal impact upon the students’ educational rights and privileges. The English proficiency of the students was intermediate or higher, with TOEIC scores ranging between 500 and 800. The students had taken a required writing course in the first semester, but had not been exposed to AWE prior to this study.

**Instruments**

The AWE software employed in this study was *ETS Criterion*. It contains grammar, usage, mechanics, style, and organization checks. For grammar, usage, mechanics, and style, students are able to see errors tallied in various bar graphs (Figure 1). For example, when students view the page for grammar, Criterion provides feedback such as the number and placement of fragments, garbled sentences, or subject-verb agreement errors. As for usage, these include such errors as wrong articles, missing or extra articles, confusing words, wrong forms of words, faulty comparisons, preposition errors, nonstandard word forms, and negation errors. Concerning mechanics, the section provides information on capitalization, missing punctuation, fused words, and duplicate words. Regarding style, Criterion can provide advice about how to revise words, phrases, and sentences so as to render these clearer and more effective. For organization, the system helps students to highlight and view the different elements of their essay, and thus students may gain a better idea of how the ideas in each of their essays relate to and cohere with one another (Figure 2).
Figure 1
Sample Page of Feedback for Grammar

Figure 2
Sample Page of Feedback for Organization
Questions have been raised by researchers about whether use of AWE feedback is pedagogically sound, suggesting that encouraging its use may be unwise. In particular, H. J. Chen et al. (2009) have suggested that the Criterion program may be essentially flawed in its system of providing error feedback. However, the researchers of this study made the decision to utilize the program because it has been the subject of previous research and other educational institutes are using it as well. Furthermore, we are not so much encouraging the use of Criterion as cautioning against an over-reliance on it, and we certainly do not wish for instructors to expect more from it than is reasonable or prudent. The fact that there are flaws in the program, however, does not mean that all of its functions are problematic. As with most educational tools, it is important to exercise discretion.

The main data-collection instrument was a questionnaire administrated in English (though students were allowed to answer the open-ended question in Chinese), which was carried out anonymously to help maintain student confidentiality and encourage them to provide fuller and more honest responses.

The first part concerned students’ responses to different kinds of feedback and was adapted from Lai (2010), who compared students’ responses to PE and AWE. In order to include perception of teacher feedback as well, we revised the items in her questionnaire to examine our students’ perceptions of all three types of feedback (that is, PE, AWE, and teacher feedback). Eleven 4-point Likert type items were included, and two open-ended questions were added to encourage comparisons of the different kinds of feedback, and to investigate the emotions students encountered while using Criterion.
The second part was the emotion indexes developed by Kay and Loverock (2008), where four emotions (anger, anxiety, happiness, and sadness) were selected on the basis of a detailed review of previous research on measuring emotion, and applied to assessing the emotions of pre-service teachers as they learned with the aid of computers. For the present study, the internal reliability of the scales and the construct validity were tested and confirmed by an exploratory factor analysis. The reliability values of the questionnaire, using Cronbach’s alpha coefficient of internal consistency, were as follows: peer = 0.90; teacher = 0.886; Criterion = 0.91. These values reached the satisfactory level, according to Nunnally and Bernstein (1994).

Finally, five students took part in an interview at the end of the course. The semi-structured interview was conducted in Chinese, and it was conversational in nature, involving open-ended questions so as to encourage the students to express their ideas more fully than might be expected in a written questionnaire. Later, the interviews were transcribed, translated and extracted to triangulate with the other data.

**Procedures**

During the first half of the semester (the nine weeks prior to the midterm exams), students worked on two pieces of paired writing. In the second half of the semester (the 9 weeks after the midterm exams) students composed two individual pieces of writing. The decision to use paired and individual writing was to encourage collaborative writing practices, and was not used as a factor in giving different kinds of feedback to students. The instructor gave each essay a final grade based on its content, organization, and grammar/mechanics.
Writing that underwent instructor-guided revisions also received an intermediate grade that focused on content and/or organization.

The first piece of paired writing underwent peer and instructor feedback and the Criterion program was not used to grade it. The assignment was a process analysis, with students being asked to write the steps to prepare a type of food or dish of their choice. The topic was consistent with the Criterion program’s topic “Directions for a Meal” (9th Grade, Topic #3).

This piece of writing included a minimum of two in-class peer reviews. Students were divided into small groups of three to four where they exchanged their writing with others. These sessions lasted for approximately 30 minutes each. The peer feedback focused on the content and organizational aspects of the writing, such as opinions regarding the introduction to the main body, which details within the main body they found most attractive, and which sections they felt needed more elucidation.

The instructor gave written feedback on the students’ printed drafts at least three times. More instructor feedback was available if the students were having difficulties or if they misunderstood the topic. The first round of feedback focused on content; the second on organization; and the third on providing grammar, mechanics, and vocabulary-related corrections. The type of feedback given by the instructors in the earlier stages—namely, content-and-organization-based—was largely similar to the feedback given by peers; the main difference was that the instructors gave more form-based feedback. However, this occurred at a later stage in the writing process, when peer feedback was not being used, and was done largely for proofreading purposes. In other words, the instructor and peer
feedback was not equivalent in thoroughness, though neither were they divergent or at odds with one another in approach.

The second piece of paired writing was completed using the Criterion program. The rhetorical pattern was persuasive/argumentative, and the Criterion program provided students with a choice of two topics: “After-School Jobs” (11th Grade / College Prep level, Topic #2) or “Guidelines” (11th Grade / College Prep level, Topic #14). Students were required to make a minimum of two submissions, but were encouraged to make more. There was no teacher or in-class peer feedback given for this writing. Out-of-class peer feedback was encouraged, but not required.

The third piece of writing, which the students did individually, was administered in a way that was consistent with the first piece of paired writing. There was peer and instructor feedback, but the Criterion program was not used. The rhetorical pattern was descriptive, and the students could choose one of two topics. These topics were based on the Criterion program topics “Private Island” (10th Grade, Topic #16) and “Your School” (10th Grade, Topic #26). Peer and instructor feedback was conducted in a manner consistent with the first piece of paired writing. That is, peer feedback was elicited during class a minimum of two times, and the instructor provided written feedback on the students’ printed drafts a minimum of three times.

The fourth piece of writing, which was the second individual writing, was done using the Criterion program. The rhetorical pattern was narrative, using the Criterion program topic “Proudest Moment” (10th Grade, Topic #17). As with the second piece of writing, students were required to make a minimum of two Criterion
submissions, and no instructor feedback or in-class peer feedback was provided.

During the last week of the course, students answered the questionnaire and participated in the interview. The researcher used a web camera to record their responses, and they were in a position to look at the AWE feedback on a computer screen at the time of the interview. Afterwards, all the data was analyzed qualitatively and quantitatively. The procedure of the study is presented in Figure 3.

Figure 3
Flow Chart of the Study
RESULTS

Students’ Perceptions of Different Kinds of Feedback

Student opinions on the various types of feedback were collected at the end of the semester. A *t*-test was employed to compare the responses of the two groups as they were taught by different teachers. The results showed that there were no significant differences (Peer \([t = -1.58, p = 0.12]\); teacher \([t = -0.92, p = 0.35]\); Criterion \([t = 1.28, p = 0.20]\)), so the main analyses were run on the combined data. Table 1 summarizes student responses to teacher feedback (T), peer feedback (P), and Criterion feedback (C). The averages (mean) and standard deviations (SD) for each category are listed under columns T, P, and C.

Next, repeated-measures one-way ANOVAs were used. The results of significant differences among the different types of writing feedback are presented in the right-most column, and the results of Mauchly’s Test of Sphericity are presented in Table 2. The Approx. Chi-Square value is 0.999 and the significance level is above 0.05. The effect sizes measured by Cohen’s *d* were \(P \text{ versus } T = 0.19\), \(P \text{ versus } C = 0.58\), and \(T \text{ versus } C = 0.79\), which show small, medium, and large effects, respectively.

The first 11 items are of the 4-point Likert type and hence a value above 2 is higher than the average: the higher the value, the more positive the perception is. Generally speaking, student responses to teacher feedback were significantly more positive than that of peer feedback; likewise, student responses to peer feedback were
### Table 1

Results of Student Responses to Different Kinds of Feedback

<table>
<thead>
<tr>
<th>Items</th>
<th>Mean</th>
<th>SD</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>1. I regard ___ as a real audience.</td>
<td>3.040</td>
<td>.656</td>
<td>.565</td>
</tr>
</tbody>
</table>
|                                                                       | 3.380| .565| .641         | T > P
|                                                                       | 2.480| .641|             | P > C
|                                                                       |      |     | T > C        |
| 2. I highly value the feedback from ___ on my writing.                 | 3.060| .608| .541         |
|                                                                       | 3.460| .541| .711         | T > P
|                                                                       | 2.350| .711|             | P > C
|                                                                       |      |     | T > C        |
| 3. I make use of ___ feedback in making revisions.                    | 3.100| .603| .505         |
|                                                                       | 3.500| .505| .690         | T > P
|                                                                       | 2.620| .690|             | P > C
|                                                                       |      |     | T > C        |
| 4. I like writing with ___ feedback.                                   | 2.960| .685| .565         |
|                                                                       | 3.380| .565| .760         | T > P
|                                                                       | 2.330| .760|             | P > C
|                                                                       |      |     | T > C        |
| 5. I revise my writing more than feedback/ ___ feedback when I use feedback. | 2.370| .627| .660         |
|                                                                       | 3.270| .660| .826         | T > P
|                                                                       | 2.440|     |             | T > C
| 6. Writing with ___ feedback has increased my confidence in my writing.| 2.900| .603| .675         |
|                                                                       | 3.230| .675| .723         | T > P
|                                                                       | 2.290|     |             | P > C
|                                                                       |      |     | T > C        |
| 7. The essay comments the ___ gives are fair.                          | 2.920| .659| .559         |
|                                                                       | 3.350| .559| .850         | T > P
|                                                                       | 2.440|     |             | P > C
|                                                                       |      |     | T > C        |
| 8. I feel the ___ won’t avoid giving negative feedback for fear of hurting the writer’s feelings. | 2.760| .710| .717         |
|                                                                       | 3.270| .717| .846         | T > P
|                                                                       | 2.900|     |             | C > P
| 9. I have enjoyed ___ feedback activities during this semester.        | 2.870| .817| .546         |
|                                                                       | 3.230| .546| .734         | T > P
|                                                                       | 2.330|     |             | P > C
|                                                                       |      |     | T > C        |
Table 2
Mauchly’s Test of Sphericity

<table>
<thead>
<tr>
<th>Within Subjects Effect</th>
<th>Mauchly’s W</th>
<th>Approx. Chi-square</th>
<th>df</th>
<th>Sig.</th>
<th>Epsilon(a)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Greenhouse-Geisser</td>
</tr>
<tr>
<td>Writing feedback</td>
<td>.979</td>
<td>.999</td>
<td>2</td>
<td>.607</td>
<td>.979</td>
</tr>
</tbody>
</table>

significantly more positive than Criterion feedback, except for items 5, 8, and 10. A detailed description of each item is presented in the following paragraphs.

Item one indicates that the students had the highest regard for their teacher as an audience, at 3.38. They respected their peers second at 3.04, and the Criterion program the least at 2.48. The students also valued the teachers’ feedback the most, as shown in item two. As reflected in item three, when asked which feedback they preferred using, students indicated they used teacher feedback the most, at 3.27, followed by peer feedback at 2.96 and the Criterion program at 2.33. The student responses demonstrate that they used teacher feedback to revise their writing more frequently than they
used peer feedback or the Criterion program’s feedback. This is clearly seen in their responses to item five. Furthermore, the students felt the teachers’ feedback helped increase their confidence level the most, as shown in item six. The students also considered the teachers’ comments to be fair, followed by peer comments and with the Criterion program’s comments coming in last yet again, as seen in item seven.

The results from item eight, on the other hand, differ from the others. When students were asked if they felt that a given feedback provider was avoiding giving negative feedback for fear of hurting the writer’s feelings, the students gave the teachers the highest average at 3.27, followed by the Criterion program at 2.90, and the peers the lowest at 2.76. This suggests that students thought their peers were holding back when giving negative feedback, due to not wanting to hurt their fellow students’ feelings. Items nine to eleven show that the students perceived the teacher feedback as the most positive of the three types: they regarded it as effective and wanted to receive teacher feedback. Peer feedback came in second in all of these items, though it was close to the ranking of the Criterion program feedback in item ten.

For every item, teacher feedback received the highest average. In most categories peer feedback was second, except for item eight where it was the lowest, and for items five and ten where peer feedback and the Criterion program had similar ratings. In other words, students preferred teacher feedback to peer feedback, and they preferred peer feedback to feedback from the Criterion program most of the time. However, students revised their writing using the
Criterion program’s feedback slightly more than they did with peer feedback.

Overall, these results are similar to previous studies comparing teacher feedback and peer feedback. For example, Zhang (1995) found that EFL students overwhelmingly preferred teacher feedback, and M. Yang, Badger, and Yu (2006) pinpointed that students were likely to adopt teacher feedback, which led to greater improvements in students’ writing. The results regarding peer feedback and computerized feedback also resemble those of Lai’s (2010) study and Fang’s (2010) study, which showed that EFL learners in Taiwan generally opted for PE over AWE. What is different from previous studies (e.g., Lai, 2010) is that students in this study claimed to adopt the Criterion program’s feedback to revise their writing slightly more than they did with peer feedback.

Responses to the open-ended question reveal one possible explanation for these results. A number of students pointed out that the primary difference between teacher and peer feedback on the one hand, and mechanized feedback on the other, is that AWE only offers one-way communication. The following two excerpts are instructive in reflecting the students’ views on this matter (All the excerpts, originally in Chinese, were translated by the authors):

My peer may not understand what I wrote, but we can discuss the confusing parts together. Criterion can only point out my mistakes and ask me to check the writing handbook. (Student C, interview)

Criterion can provide us with timely feedback, but sometimes it cannot give more advice like the teachers do. (Student B, open-ended question)
Students can only receive the AWE feedback passively, and they do not have an opportunity to clarify their doubts regarding the best way to make use of the feedback. As a result, when they are confused by the feedback, they can only guess at the possible reasons for it. On the other hand, they can seek out their teacher or peer for a more sophisticated explanation of problems found in their writing. This explains why the students generally prefer human feedback to machine feedback, as “writing is a social-communicative act involving the negotiation of meaning between writers and readers” (C.-F. Chen & Cheng, 2008, p. 109) and that “learning to write well is a demanding and insuperable task, it may be helpful to help students recognize that they are not alone in their anxiety” (Cheng, 2004, p. 57).

Students did not appreciate peer feedback as much as teacher feedback, because they felt that peers tried their best to avoid giving negative feedback in order to reduce socially uncomfortable situations—a finding consistent with much of the literature (e.g., Cross & Hitchcock, 2006; Hu & Lam, 2010; Lai, 2010). Furthermore, students were sometimes not confident regarding the reliability and validity of peer feedback, due to variations in their peers’ language proficiency (Min, 2006). Excerpt 3 illustrates this perspective:

Peer feedback is restricted by the peer’s English competence. A peer with high English proficiency can tell you where improvements are needed. But up to now, most of my peers have failed to make suggestions about structures and organization. I think they are not as good as the teachers in terms of their English proficiency. (Student E; interview)
In a similar vein, Hu and Lam (2010) indicated that L2 learners’ limited knowledge of the target language and its rhetorical conventions, lack of experience with and training in the use of peer review, and a complex of cultural and social differences all related to problems encountered during peer review.

Conversely, other responses show a relatively positive attitude towards AWE feedback. Excerpts 4 and 5 show the perceived usefulness of AWE:

Criterion can provide me with immediate feedback on grammar and other mistakes. It is really helpful for revising my writing. (Student D, open-ended question)

I like the immediate feedback from the AWE. It may be not as good as teacher or peer feedback, but I can use it anytime I get access online. (Student E, interview)

The above results echo the findings of previous studies which show that AWE feedback is perceived as being helpful in improving EFL college learners’ writing proficiency because of the immediacy of the feedback (N. D. Yang, 2004) and the fact that the feedback is perceived to be helpful for revisions (Fang, 2010; Yeh, Liou, & Yu, 2007). This study demonstrates that the ubiquitous character of AWE feedback is another reason why students regard AWE as useful, because they may not always be able to get help from teachers and peers.
The Emotions Arising from the Use of the AWE Tool

Students were also asked about how they felt when they were learning how to use Criterion. The emotions students experienced in their use of the Criterion program were divided into four major categories: happiness, sadness, anxiety, and anger. In Table 3, the more detailed emotions that students might have experienced are listed below each category in the table. The options for each category

Table 3
Emotions Evolved in the Writing Process Using Criterion

<table>
<thead>
<tr>
<th>Emotions</th>
<th>M</th>
<th>SD</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happiness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Satisfied</td>
<td>2.52</td>
<td>.671</td>
<td></td>
</tr>
<tr>
<td>5. Excited</td>
<td>2.31</td>
<td>.673</td>
<td>2.46</td>
</tr>
<tr>
<td>9. Curious</td>
<td>2.66</td>
<td>.745</td>
<td></td>
</tr>
<tr>
<td>Sadness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Disheartened</td>
<td>2.63</td>
<td>.627</td>
<td>2.53</td>
</tr>
<tr>
<td>6. Dispirited</td>
<td>2.47</td>
<td>.703</td>
<td></td>
</tr>
<tr>
<td>Anxiety</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Anxious</td>
<td>2.54</td>
<td>.699</td>
<td></td>
</tr>
<tr>
<td>7. Insecure</td>
<td>2.33</td>
<td>.653</td>
<td>2.34</td>
</tr>
<tr>
<td>10. Helpless</td>
<td>2.41</td>
<td>.804</td>
<td></td>
</tr>
<tr>
<td>11. Nervous</td>
<td>2.22</td>
<td>.673</td>
<td></td>
</tr>
<tr>
<td>Anger</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Irritable</td>
<td>2.51</td>
<td>.644</td>
<td></td>
</tr>
<tr>
<td>8. Frustrated</td>
<td>2.62</td>
<td>.718</td>
<td>2.46</td>
</tr>
<tr>
<td>12. Angry</td>
<td>2.33</td>
<td>.879</td>
<td></td>
</tr>
</tbody>
</table>

Note. 1 = None of the time, 2 = Some of the time, 3 = Most of the time, 4 = All of the time.
are (1) none of the time, (2) some of the time, (3) most of the time, and (4) all of the time. Therefore, a value above 2 is higher than the average.

The results of descriptive statistics, including the mean and standard deviations, are presented in Table 3. Regarding the results of different categories, the happiness category generated an average of 2.46 (containing three sub-items: satisfied = 2.52, excited = 2.31, and curious = 2.66). The second category, sadness, has the highest score among the four categories, with an average of 2.53 (disheartened = 2.63, and dispirited = 2.47). The third category is anxiety, which has an average of 2.34 (anxious = 2.54, insecure = 2.33, helpless = 2.41, and nervous = 2.22). This is the lowest value among the four categories. The final category, anger, generated an average of 2.46 points, the same as the happiness category (irritable = 2.51, frustrated = 2.62, and angry = 2.33).

Next, the results of Mauchly’s Test of Sphericity, which tests for one of the assumptions of the ANOVA with repeated measures, are presented in Table 4. The Approx. Chi-Square value is 22.115 and the significance level is below 0.05. As a result, we proceeded with the test using the correctional adjustment of Greenhouse-Geisser. The results of tests of within-subjects effects show that when using an ANOVA with repeated measures with the Greenhouse-Geisser correction, the mean scores for the four constructs of emotions were not statistically significantly different ($F(2.244, 114.463) = 1.309, p > 0.05$). Eta square ($\eta^2$) for effect size was 0.025.
Table 4

Mauchly’s Test of Sphericity

<table>
<thead>
<tr>
<th>Within Subjects Effect</th>
<th>Mauchly’s W</th>
<th>Approx. Chi-square</th>
<th>df</th>
<th>Sig.</th>
<th>Epsilon(a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emotion</td>
<td>.623</td>
<td>23.538</td>
<td>5</td>
<td>.000</td>
<td>.748</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.784</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.333</td>
</tr>
</tbody>
</table>

The non-significant results mean that the different types of emotion were equally strong or prevalent. Diverging from some previous studies (e.g., Lai, 2010), the construct “anxiety”—which has tended to be the most frequently discussed emotion in CALL—was not the primary emotion experienced by the students in this study. From the interviews with five students, we found that despite their initial curiosity and excitement, students in this study felt impatient and confused when they received repetitive and vague AWE feedback, believing it to be unhelpful to the revision process. The following views make this clear:

I am curious about the feedback that a computer can give us. After more practice, I know how to use Criterion, so I don’t feel anxious now. But I feel impatient because sometimes I don’t know how to revise my writing, since the feedback is so vague. I don’t know how to describe my feelings. Unhappy and impatient! (Student A, interview)

“Oh my God, so many mistakes!” This is how I felt when I saw the AWE feedback. I am really too confused to revise my writing. (Student F, open-ended question).
The above excerpts show that due to the students’ unfamiliarity with the definitions of some writing feedback, such as the difference between a thesis statement and topic sentence, and their positions in an essay, students remarked that the software was not giving enough hints or was not specific enough. For instance, student F needed to spend time defining the jargon that confused her.

Some students, however, indicated that despite the disadvantages of mechanized feedback, the Criterion program is available for use whenever necessary, whereas teachers might be tired or busy when students need help. In other words, students were open to using machine feedback, especially that which was related to form.

Most of the students expressed views in the interview to the effect that the software is helpful on systematic formatting but not on giving advice for further revisions of the essay. In addition, according to the interview, most students did not feel anxious about using Criterion. One student mentioned that since the scores on Criterion were only made available to the students themselves, and other students would not be able to see them, it would not influence their emotions. As a result they would not feel anxious. Furthermore, Criterion was seen as offering more objective feedback, not susceptible to rater subjectivity. The following excerpt illustrates both of these points:

I feel nervous about teacher feedback because different teachers have different standards, but I feel comfortable about AWE because no one else can see the feedback. (Student B, interview)
Moreover, as the students became more familiar with the interface, they grew pleased as the number of errors decreased. Student D, in particular, recommended this software as the following excerpt shows.

> Teachers may get tired after they correct so much writing, but the machine won’t miss any mistake when we use it individually. (Student D, interview)

All the students agreed that if the software offered more hints and advice, it would be more helpful. Finally, one of the students had the following insightful comment.

> I have to spend time and effort in making sense of the feedback and then try to revise the draft. I believe that whether the student has problem-solving ability and independent thinking skills is crucial for the negative emotions aroused by the AWE software. (Student A, interview)

The above excerpt echoes a finding which other recent studies (e.g., Järvelä, Veermans, & Leinonen, 2008; Järvenoja & Järvelä, 2005; Volet & Järvelä, 2001) have pointed out, which is related to students’ emotional and motivational experiences while taking part in computer supported learning projects. That is, students with differing socioemotional orientation tendencies interpret the new instructional designs in ways which subsequently lead to different actual behaviors (Järvelä, Lehtinen, & Salonen, 2000). Järvelä, Hurme, and Järvenoja (2010) indicate that students may adopt context-specific interpretations of motivational goals and self-regulation when confronted with atypical learning demands. In our context, in addition to giving writing feedback, the AWE software asked students to refer
to a writing handbook. If the students lacked motivation or independence in learning, they might easily give up on making the needed effort to improve their writing.

To sum up, students’ emotional reactions to academic tasks are related to student motivation, learning strategies, and self-regulation, as Pekrun, Goetz, Titz, and Perry (2002) have pointed out. Similarly, Wosnitza and Volet’s (2005) study highlighted the contribution of students’ reflections on their emotions in facilitating the learning process. They also pointed out that one challenge in studying emotions is the limited access to the affective reaction process, which is difficult to overcome unless learners are willing to disclose their emotions. Students in this study reflected on the learning process at the end of the course, which may partially explain why they reported similar feelings for all kinds of emotions. Another possible explanation for this phenomenon is the context of blended learning: namely, the students in this study also used teacher feedback and peer feedback. Although students were asked to reflect on the emotions they felt when using the Criterion program, it might have been difficult for them to differentiate the sources of their feelings.

CONCLUSIONS AND SUGGESTIONS

This study aimed to explore student responses to different types of feedback for writing, as well as the unintended consequences of the use of automated writing evaluation. The primary findings can be delineated into three main observations. First, students consistently preferred teacher feedback to either peer or AWE feedback, suggesting that the teacher’s role in the writing classroom is still
viewed as central. Second, although students tended to prefer peer feedback to AWE feedback, they nonetheless perceived the potential shortcomings of peer feedback, most particularly regarding the willingness—or lack thereof—of peers to provide feedback which might be perceived as negative or critical. However, they adopted AWE feedback more frequently than peer feedback when making revisions, and regarded the ubiquitous feature of AWE as helpful in their autonomous learning. Finally, students did not experience writing “anxiety” any more strongly than they experienced other emotions. If anything, they expressed at least equal levels of positive curiosity and of being disheartened as they did of anxiety.

The limitations of this study can be summarized as follows. First, this study involved two writing classes, taught by different instructors. Although the teachers of each class followed the same curriculum, differences in teaching styles are almost certainly inevitable, and so it could be argued that these differences could not be completely ruled out as a factor which could affect the results. However, in order to minimize the likelihood of this, a $t$-test was run on the data from the two classes, and no significant differences were found.

Second, only self-reported emotions based on the questionnaire were investigated. Actual behavior or other emotions experienced while using the software were not examined. Still, within this limitation were steps taken to provide triangulation. Since this study is predominantly a quantitative exploration of the students’ affective responses to an AWE system in English writing feedback, the researchers included an open-ended question in the questionnaire and a small-scale follow-up interview to probe more in-depth reflections.
of the students’ emotions. It is hoped that this has helped to provide a fuller, more nuanced analysis, if not a complete analysis.

The implications of the study are as follows. First, teachers who hope to increase the use of peer and/or AWE feedback in the writing classroom need to pay more attention to how to raise the profile of each type of feedback. For instance, instead of expecting peer and AWE feedback to meet with students’ approval in all areas, it might be more realistic to focus on areas where that feedback is felt to be more helpful, such as using AWE feedback to address certain form-related areas. As C.-F. Chen and Cheng (2008) pinpointed, despite preferences for teacher feedback, the ways in which teachers integrate AWE into the writing class can influence student perception of AWE and help make them more accepting of its contribution. Second, peer feedback may require more coaching by teachers. In particular, students need help in giving constructive criticism to their peers in a manner that does not come across as overly negative. Hu and Lam (2010) also emphasized that certain training procedures should be provided to help students increase the provision and incorporation of valid suggestions addressing peer feedback. Otherwise, students will tend to be reticent in offering critical feedback. Finally, teachers need to beware of exaggerating the role played by any single affective reaction in the classroom. Teachers should better understand the complicated emotions which arise in students during the process of computer assisted learning so as to facilitate effective learning outcomes. In cases where there is a complex interplay of positive and negative emotions—such as curiosity and sadness in this study—awareness can aid teachers who want to build on the strengths of positive emotions and mitigate the effects of negative emotions.
Based on the findings of this study, there are also some suggestions for software designers. The assistance and assessment functions of AWE are still being developed and have not yet reached maturity (C.-F. Chen & Cheng, 2008; H. J. Chen et al., 2009). Indeed, students in this study expressed negative feelings towards the vagueness of the machine feedback. Consequently, more effort should be spent on improving the reliability and validity of the scoring process. Moreover, students in this study reported feeling disheartened and dispirited when using Criterion. Software designers, therefore, should consider how the instructional and user-interface designs of AWE help students. Specifically, perhaps designers could consider taking more cues from the types of technological devices and applications that students make use of in their daily lives, so as to ease the transition into using these tools in an educational setting and to squarely address the motivation issues. If such changes were made, students would be in a much better position to make use of the AWE software, especially in their individual writing. In this way, AWE could better promote learner autonomy in EFL writing (Benson, 1997, 2001).

Suggestions for future studies include a more in-depth review of emotions which arise during the process of computer assisted learning. Obtaining contemporaneous information will require identifying how to access students’ emotional reactions as they write and receive feedback. Weekly journals, for example, could be adopted to collect information about the emotions involved in a specific writing activity and their sources. Future studies could also focus on the timing of teacher intervention (Wosnitza & Volet, 2005) after identifying students’ negative affective reactions. Timely adjustment
of teaching strategies and appropriate scaffolding could help students conquer their negative affective reactions and better allow them to benefit from computer-assisted learning. Last but not least, due to the anonymous nature of the survey, we did not explore the relationship between learner emotion and performance. Relevant studies should be conducted to reveal how emotions influence the learning process and learning outcomes so that teachers can learn how to adopt software for EFL writing in an optimal way.

ACKNOWLEDGMENT

This paper is partially supported by a grant from the National Science Council in Taiwan (NSC 100-2511-S-328-001). The authors would like to thank the students who participated in this project, and of course special thanks are in order for the reviewers and the editor for their insightful comments and valuable feedback.

REFERENCES


ABOUT THE AUTHORS

Mei-jung Wang is an associate professor at the Department of Applied English of National Kaohsiung University of Hospitality and Tourism, Taiwan. Her current research interests include literacy instruction, computer assisted language learning, and English for Specific Purposes.

David Goodman is a senior lecturer at the Department of Applied English of National Kaohsiung University of Hospitality and Tourism and a PhD candidate at National Kaohsiung Normal University. His interests cover linguistics and EFL writing.
寫作自動回饋系統：學生的感知和情緒感受

摘要
現代教育科技的發展使得個別化學習及寫作自動回饋系統深深吸引語文老師。過去對於寫作自動回饋系統之研究多聚焦於評分的效度、老師的教學法、或學生感受的成效。關於寫作自動回饋系統造成學生的情意反應如較少受到關注。本研究探討大學寫作課中學生對不同之回饋的感知及使用寫作自動回饋系統時，學生所感受的情緒。共有四十六位修習寫作課的大學生參與。本研究比較學生對教師回饋、同儕回饋、及寫作自動回饋的反應及使用寫作自動回饋系統時感受之情緒。研究發現可提供教師如何有效將寫作自動回饋系統融入教學之參考。

關鍵詞：寫作回饋　寫作自動回饋系統　情緒