Computer Assisted Language Instruction Consortium

CALICO JOURNAL (ONLINE) ISSN 2056–9017

Article

Perceptions of L1 Glossed Feedback in Automated Writing Evaluation: A Case Study

Jayme Lynn Wilken

Abstract

Learner perceptions toward and utilization of L1 glossed feedback in an automated writing evaluation (AWE) program were investigated in an Intensive English Program (IEP) class. This small case study focused on two Chinese students who responded to weekly surveys, semi-structured interviews, and screen capture videos of their revisions over a four-week period. In weeks 1 and 3, the students received English-only feedback (L2), and in weeks 2 and 4, the students also received feedback in their native language (L1). The data were recorded, transcribed, and coded. Because the L1 has been shown to be helpful in students' learning, it was hoped that the L1 glossed feedback in AWE would prove helpful as well. The participants felt a need for the glosses but also expressed reservations about relying on the L1. While the participants' revision behaviors sometimes differed, both showed a positive attitude toward the L1 glossed feedback, toward increased noticing of errors, and toward their autonomy while using AWE.

KEYWORDS: AUTOMATED WRITING EVALUATION, AUTONOMY, L1 GLOSSED FEEDBACK, PERCEPTIONS

Introduction

Automated writing evaluation (AWE) enables students to receive feedback on a piece of writing within seconds after submission. With the click of a mouse, a student can receive holistic scores, reports detailing categorized errors, and feedback on those errors. Over the past decade, AWE research seems to be slowly shifting toward perceptions (e.g., Chen & Cheng, 2008;

Affiliation

Iowa State University.
email: jaymew@iastate.edu

CALICO JOURNAL VOL 35.1 2018 30–48

https://doi.org/10.1558/cj.26383

Link, Dursun, Karakaya, & Hegelheimer, 2014) and effective use of this and other computer assisted language learning (CALL) tools (e.g., Hegelheimer & Tower, 2004).

Since a learner may need to have comprehensible input for effective learning to occur (Krashen, 1985), modifications may be one way to make input comprehensible "for acquisition" (Larsen-Freeman & Long, 1991, p. 144). These modifications can include glosses in AWE, where learners can request a response from the computer about what they have just written. Learners can then respond to that feedback by revising their essays, potentially producing "comprehensible output" (Swain, 1995) as a result. AWE can now provide L1 glossed feedback, i.e. feedback that appears in the students' native language, as a modification to its English feedback.

Even though research has shown L1 glossed feedback to be helpful in areas such as vocabulary acquisition (e.g., Prince, 1995; Laufer & Shmueli, 1997), the area of L1 glossed feedback has not yet been studied in conjunction with AWE. The present study investigated how students perceive and utilize the resources available in Criterion, an AWE program by Educational Testing Service (ETS), when feedback is available in their native language in addition to English.

L1 Glossed Feedback in AWE: A Multifaceted Issue

Depending on the context in which AWE is used, students may react quite differently to it and may even be unwilling to use it, due to their negative perceptions (Attali, 2004; Chen & Cheng, 2008; Li, Link, & Hegelheimer, 2015). However, used in conjunction with human feedback, AWE may be useful in providing necessary scaffolding for learners (Cotos, 2011; Hyland & Hyland, 2006). Further, one effective way to enhance this scaffolding may be through providing feedback in the learners' first language (L1).

L1 Use in the Classroom

With the rise of communicative teaching methods, the L1 has been neglected, and sometimes even banned, in the classroom context (Cook, 2001). Even if it is not banned outright, the L1 is likely not being utilized as the resource that it could be (Atkinson, 1987; Cook, 2001), though research has found the L1 to be a helpful tool for providing scaffolding, maintaining interest, developing management strategies (Antón & DiCamilla, 1998), formulating goals (Brooks & Donato, 1994), promoting joint understanding, focusing attention on specific items, and moving tasks forward (Swain & Lapkin, 2000). Not all research has shown such strong favor for L1 glosses, however. Some studies (Jacobs, Dufon, & Fong, 1994; Yoshii, 2006) have shown no significant difference between the L1 and L2 gloss conditions regarding vocabulary acquisition,

where text + images seemed to produce the most significant results, regardless of the language used.

Merely providing L1 glosses, however, will not aid students' acquisition if they do not utilize them. For glossing of any type to be helpful, students need to be motivated enough to access them, which they often choose not to do (e.g., Hegelheimer & Tower, 2004). Furthermore, Hulstijn, Hollander, and Greidanus (1996) argue that learners should be provided with "easily accessible glosses and learner-friendly, non-tedious review opportunities" in order to "follow up on incidental learning with intentional learning" (p. 337). To encourage their use, then, glosses should be intentionally introduced (Hegelheimer & Tower, 2004) to help students recognize the glosses, understand their options, and evaluate optimal timing of their use (Hubbard, 2004).

Errors and Noticing of Glosses

Since learners need to notice, attend to, and revise their errors, making errors salient may contribute to moving learners down the path toward acquisition (Cotos, 2011; Schmidt, 1990). Because a final factor in this revision progression may be the learners' attitudes and beliefs about the feedback they are receiving (Storch & Wigglesworth, 2010), it seems imperative for educators to have a clear understanding of how to provide corrective feedback (CF) effectively to help students achieve their goals.

In an IEP, students' goals often include passing high-stakes entrance tests, so strengthening basic essay format and grammar skills are of interest to them. AWE seems that it may mesh well with this goal; however, because student motivation in accessing the available glosses is of concern, learner perceptions of the glosses in AWE should be investigated. In fact, Levy and Stockwell (2006) emphasize, "One cannot ignore students' attitudes toward the tasks they are asked to complete, because there is every likelihood that their attitudes will affect their performance" (p. 174).

In response, the present case study collected data on learner perceptions of L1 glossed feedback in AWE through surveys, interviews, and screen recordings in an attempt to view the writing and revising experience through students' eyes.

Research Questions

To guide the investigation, the following questions were asked:

1. What are Intensive English Program (IEP) students' perceptions of the usefulness of the L1 glossed feedback?
2. How do IEP students make use of the L1 glossed feedback themselves?

Method

Participants

The participants, two Chinese males, were enrolled full-time in an Intensive English Program at a large Midwestern university during the fall semester of 2012. They had tested into a low-intermediate writing class (level 3 out of 6) with nine other students by taking the Michigan Test and in-house placement and diagnostic tests. Even though the class as a whole (n = 11) participated in the in-class AWE activities (Wilken, 2013), this study focuses on two 18 to 20-year-old Chinese male volunteers, Stephen and Amos (names anonymized) who participated in in-class and out-of-class activities. Both had studied English for 12 years and had written in English for 10 years, though neither had used AWE before this class. They also signed Institutional Research Board (IRB) releases signifying their agreement to take part in weekly interviews and surveys, in addition to screen capturing of their revisions with AWE.

Data Sources

Data were taken from: (1) the students' data saved by Criterion (holistic scores, submissions, word counts, and time spent); (2) the videos of revisions, recorded in a screen capturing program, Camtasia by TechSmith, and exported as mp4 files; scripts were written from these recordings and relevant audio was also noted in the scripts; (3) the complete transcripts of the interviews; (4) the results of the weekly surveys, a mix of Likert scale and short answer questions; and (5) a brief bio-data questionnaire.

The Task

Over a four-week period, the students were asked to respond weekly to writing prompts and to revise essays with either English or bilingual feedback; furthermore, they shared perspectives from this experience through surveys and interviews.

The task was designed with both the learners' and class objectives in mind. Specifically, during a 50-minute class each Monday, students wrote a three to five paragraph essay from a prompt which was chosen by the teacher from a list of Test of English as a Foreign Language (TOEFL) prompts Criterion provides. Revising essays was their task during lab days on Wednesdays; Thursdays, at the beginning of class, they were asked to complete a brief survey (< 10 minutes) about their weekly experience with AWE. Finally, the focus participants came for a 15-minute semi-structured interview each Friday in the teacher's office, with questions based on their survey answers and on scripts produced from their Camtasia recordings. Over a two-week period, the students were trained to use Criterion with training videos and materials, developed by the

university and by ETS. The students also participated in two cycles of writing and revising short essays before data collection began.

On writing days, students would sign in to the program, read the prompt, and begin their response. On revision days, the students would work as usual while Camtasia ran in the background of Amos and Stephen's computers. As students worked on their AWE revisions, the teacher would walk around the class, taking the time afforded her to conference with students about their essays, to make suggestions, and to offer assistance.

Criterion, meanwhile, was also providing them with several layers of feedback, which students could access within the revision screen (Figure 1).

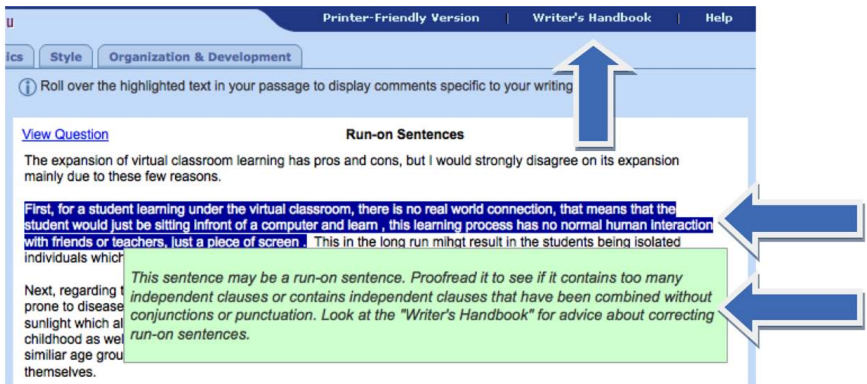

Figure 1: Feedback types within Criterion.

As a student enters the revision screen, errors can be seen highlighted in blue, and students may choose to revise based solely on this visual feedback. However, if students want to know more about an error, they can take two additional actions: (1) they can move the cursor over the blue highlights to reveal green feedback boxes which pop up on rollover. These boxes provide written feedback about the error. If this is not enough explanation, (2) the student can click on the Writer's Handbook (written feedback) and scroll down through explanations and examples of similar errors.

Criterion offers feedback in five languages: English, Japanese, Korean, Simplified Chinese, and Spanish (www.ets.org/criterion/about/). The L1 and L2 feedback appears on rollover of the blue highlighted errors (see Figure 2).

The L1 glossed feedback will also appear on click when one enters the Writer's Handbook, a feature giving additional explanation and examples of the pertinent error.

The students had access to the L1 glossed feedback on weeks 2 and 4. This pattern was designed to allow learners to compare their experiences of using

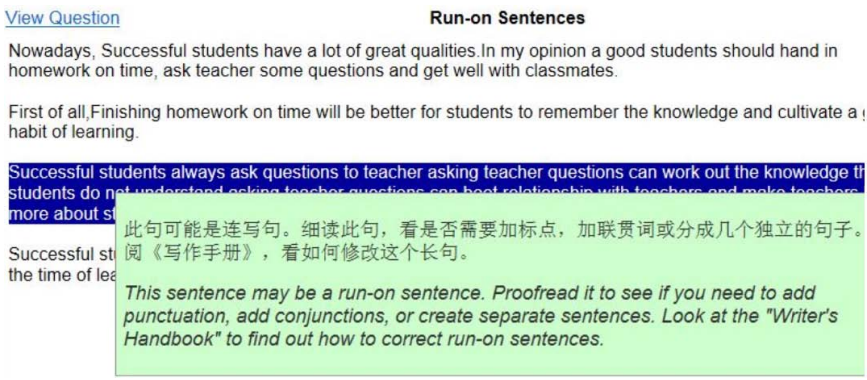

Figure 2: L1 glossed feedback within Criterion.

English-only feedback versus English + L1 feedback. Because the teacher would be there to assist the students, this schedule did not seem like it would be a hardship for the class to follow.

Analysis

Data were collected, surveys were amended, interviews were conducted and transcribed (Goh, 2002), and screen capture scripts were produced. Open coding was used to analyze the interview transcripts (Esterberg, 2002), and five coding themes from those interviews were identified: preferred types of feedback, preferred amounts of feedback, beliefs about L1 glossed feedback, comments about autonomy, and comments noting changes in perspectives.

Inter-rater reliability was calculated for both the interview transcripts themselves and for the coding of the transcriptions. Near unanimous agreement was reached for the interview transcriptions between the researcher and a Ph.D. candidate in linguistics, with the second-rater checking approximately 20% of the transcription. Where there was a discrepancy, the tape was played and agreement was reached. Twenty-five percent of the interview coding was also checked, fulfilling the suggested amounts by Mackey and Gass (2005). Inter-rater reliability was calculated to be 0.94 for the coding, a sufficiently high level of agreement.

The screen capture videos were watched and scripts were produced each week, listing times and activities occurring at those times. Audio notations in the scripts were used to answer RQ2: What did the student actually do in response to the feedback given by AWE? The scripts were also used to formulate questions for the interviews. Finally, three native Chinese Ph.D. linguistics students who work with AWE reviewed the simplified Chinese feedback in Criterion, due to comments in the pilot study about some of the L1 words

being awkward. The L1 feedback was reviewed in four areas: terminology (difficulty of words/word choice), grammar, fluency/flow, and mechanics. On a 5-point scale, with 5 being "very clear," 3 being "somewhat clear," and 1 being "not clear," seven ratings of 3 or less were given in the "difficulty of words" category (with 5 being "very easy for an IEP student to understand," and 1 being "very difficult"). Further, word choice was marked four times with a 3 or less, with 5 being "very accurate" and 1 being "very frequent errors." One reviewer explained, "The translation is kind of awkward to me because some translated phrases, such as 'proofread' are not commonly used in Chinese." Overall, though, the ratings for the L1 translations were positive with 44 out of 55 ratings (80%) receiving a 4 or 5.

Results and Discussion

Students' Perceptions of the Usefulness of the L1 Glossed Feedback (RQ1)

Stephen's Initial Reactions

While Stephen has had 12 years of English study, his speaking was often labored as he searched for the right words. On the first weekly survey, he gave his confidence in English a 2 out of 4 ("somewhat disagree") and described his current ability to revise essays by saying, "Just by myself, just by myself, maybe I can look at something obvious mistake, but I still need a teacher to help" (week 1, interview). He also compared his English skills to the other students' abilities in the class, saying "my English is not as good other, I cannot concentrate on writing lots of times, so for me, I will want to just concentrate, use my energy, concentrate on the point I need" (week 2, interview). Stephen did appear to concentrate and work carefully on his essays, and after the two training weeks, he reported having positive feelings for Criterion and predicted, "After I know the function of this software, I feel … It will help me to figure out lots of problem that I usually make" (week 1, interview). The next week, when Stephen discovered the translations by rolling over the blue highlights at the beginning of the class, he exclaimed, "Oh, teacher! How did you do that? You did an amazing thing … really helpful! Now I understand this!" (week 2, Camtasia).

Amos's Initial Reactions

Amos, the other participant, also studied English for 12 years and functioned well in English, though sometimes his listening ability seemed to cause miscommunication. He always sat in the front of the room and was much more social than Stephen, often talking to friends before, during, and after class. When Amos was asked about his confidence in correcting his own essays, he rated his confidence as a 3 ("somewhat agree") on a Likert scale survey

question and explained why he did not fully agree (4): "Yeah, I just I don't choose four because I just have problems in some academic words and some … sentence. I'm still Chinglish English. Is not very, so just a three" (Amos interview, week 1). Here, he showed awareness of his shortcomings and even used a slang term, "Chinglish," to describe his own perceived level of English competence.

When Amos spoke of his initial reaction to seeing the Chinese translations, he said, "I can know what is my wrong. I can, in the first time, I can correct the answer" (week 2, interview). He seems to be reflecting here on gaining a sense of hope after having used the L1 translations, and this high positivity toward Criterion carried throughout the study. His feeling of being "lucky" (Amos, week 3, interview) to use AWE sets the stage for many of his responses, but his specific comments and behaviors do not always match with this exuberance, as will be seen in the discussion of research question 2. Overall, however, both subjects felt positively toward their initial experience with Criterion.

Beliefs about the L1 Glossed Feedback's Clarity and Ease of Use

The issue of difficulty in understanding the translations did surface in the interviews. Stephen explained that the translations were "hard because the Chinese especially is so different from the American words" (week 3, interview). Amos explained the difficulty by saying, "The Chinese words is not hard words, but the meaning is very complex … the words is easy, but the sentence meaning is … not easy to understand" (week 4, interview). Both explained that the structure of the languages and the difficult content is what makes the L1 translations complex, as the raters also mentioned (see Analysis). Even with these critiques, the students did not seem overly disturbed by this, and did not attribute their amount of L1 use with the occasional awkwardness of the translations.

Both students preferred the visual feedback (blue highlights) calling it "easy to understand" (Stephen, week 1, interview), "obvious," and "convenient" (Amos, week 1, interview). They felt the written feedback (green box feedback/Writer's Handbook) was sometimes difficult and vague, also noted in research (e.g., Chen & Cheng, 2008).

Beliefs about Appropriate Use of the L1 Glossed Feedback

When asked to opine on the value of the L1 translations for differing levels of English proficiency, both, perhaps predictably, believe low-level students need the native-language feedback the most. In the final week, Stephen remarked, "Ya, I think that low level is most need … because maybe they can't understand that, even master grammar very well, so they need their native language help them to understand" (week 4, interview). Research bears this out:

lower proficiency students can benefit from using the L1 in complex tasks (e.g., Storch & Wigglesworth, 2012).

While Stephen and Amos both used L1 translations (in both Criterion and bilingual dictionaries), their opinions were mixed on whether higher level students would access the L1 glossed feedback. Stephen stated, "Students always lazy … even the high level student, if there is easy way to fix theys mistake, I think they will choose to use the native language to help them understand" (week 4, interview). In relation to himself, Stephen saw the possibility of L1 glossed feedback becoming less necessary:

> STUDENT: So now I'm familiar with the run-on sentence, so now if I see the green box, I can understand that what's the mistake I make … after familiar, it is not needed, the Chinese anymore.
> TEACHER: Ok.
> STUDENT: But not still [laughing]!
>
> (week 4, interview)

Even at the study's end, Stephen wanted to make sure that it was understood that he was not yet ready for L2 feedback only; he still felt that he needed the Chinese translations. However, Amos seemed to feel that upper-level students should not use the L1 frequently, and his own desire was to not rely on his L1:

> STUDENT: I think English more useful, this here, because only in this way, we can change the idea, don't [change] the thinking-style because [it is] in Chinese.
> TEACHER: You want to start thinking in English?
> STUDENT: English, English, not translating into Chinese, then come back.
>
> (Amos, week 4, interview)

Several times, Amos expressed his desire to not translate back and forth between the two languages and to not rely on the L1 for any longer than is necessary.

Beliefs about the Benefits of Glossed Feedback

The comments about the benefits centered on their reported increase in autonomy and in noticing of errors.

In Relation to Autonomy

Both students felt that Criterion, in general, helped advance their autonomy in writing. When Stephen was asked in week 4 whether he believed Criterion fostered his independence, he answered positively:

> Because Criterion can make us to write the essays more independently, and because if we just writing by myself and ask the teacher, they will waste lots of time, and maybe you can't find the teacher sometimes, and Criterion is more freedom.
>
> (week 4, interview)

This touches on an oft-discussed concern that writing teachers do not have enough time to provide quality feedback on multiple drafts (e.g., Grimes & Warschauer, 2010). If this concern causes students to leave questions unasked, their language development could be inhibited. In this case, at least, Stephen seems to be finding "freedom" from these concerns through AWE. More specific to our discussion of L1 feedback, Stephen felt revisions with the L1 glossed feedback were "more" and "quicker" (week 2, interview), a particular relief, it seemed, to a student who reported struggling with patience (e.g., week 1, interview). At this proficiency level, if a student finds the feedback to be too hard to understand, he may not access it, no matter how potentially helpful it may be.

Amos, explaining his agreement that Criterion's feedback helped his writing improve, stated, "Ya, because it let me know what errors I always do. Fix that" (week 4, interview). The academic terms, in particular, were noted as helpful in the L1. For example, after the first L1 week, Amos said, "This time it's the Chinese, so I can understand 'fragment' [smiling]" (week 2, interview), a word he had not previously understood. When discussing the text-heavy Writer's Handbook after an English-only week, he stated, "if we have Chinese translation, maybe we will use more Writer's Handbook" (week 3, interview).

Although these comments were positive toward the L1 glossed feedback, the students both clearly valued human feedback. Stephen referenced his trust in teacher feedback by saying, "maybe the teacher can explain more well than just the box" (week 2, interview). Amos was even stronger in his opinion, saying, "We can ask you [the teacher] and many friends to understand, so I don't think the native language, ya [TEACHER: made a difference?] [He nods]" (week 4, interview). Because Amos often relied on human interaction (teacher, peers, self) to complete his tasks, he may not have felt as great a need for the L1 feedback in AWE as Stephen did, who, although he valued teacher feedback, also shared reservations about the challenges of reliance on human feedback.

Regardless of which type of feedback the students report preferring, human or AWE, the main concern is that the students actually use the feedback they are given to become more proficient and self-sufficient.

In Relation to Noticing

Stephen and Amos agreed that AWE has helped them notice their most frequent errors. Stephen explained, "Criterion can tell you the popular mistake, the mistake you always make and can help you to fix [them]" (week 4, interview). Amos, likewise, believes the program can help him notice and reduce his mistakes:

> [Criterion] can give us, give me, some information about what mistakes I always do, and I think sometimes later I can remember and to correct, don't mistake in the future … the beginning, 6 or 8 [mistakes], but this time it's a 5 or 6 and when I fix it, I think it's better.
>
> (Amos, week 1, interview)

In week 4, Amos's interview revealed an example of his remembering and correcting. He was speaking about his possible future use of AWE, when Amos caught himself making a grammar error; he self-corrected and moved on. The sentence was: "For in the future, we have many essays we need to give to the professor, so we can use that to make our essays more, no 'more,' just better" (week 4, interview). Criterion flagged this very mistake in his essay this week, "more better," and he had fixed it correctly (week 4, Camtasia); now, in the interview, he corrected this same mistake in his spoken language. This could be regarded as an example of corrective feedback impacting future output, something Truscott and Hsu (2008) doubted happening as a result of CF. In this example, it seems that the AWE feedback may have indeed led to future learning.

While the surveys and interviews were helpful in understanding students' reactions to the L1 feedback they received, it is important to look at what the students actually did with the feedback they received as well (Warschauer & Ware, 2006).

Students' Use of L1 Glossed Feedback (RQ2)

To investigate this question, data from two sources were used: (1) Criterion (holistic scores, submission numbers, word counts, and time spent on each essay), and (2) Camtasia scripts (videos + audio), which recorded the types of feedback they accessed, the time spent with them, any audible questions the students may have asked the teacher or other students, and whether they accessed other online tools.

Even a cursory glance at these students' Criterion data over the four-week study, stored in the program itself, shows rather high similarity (Table 1).

Stephen and Amos's holistic scores were virtually identical throughout the study, differing by only one point. Each submitted nearly 25 times. Attali (2004) looked at submission rates and found a "general linear increase … with increasing submissions" (p. 18), though here, they both received "6"

Table 1
Students' Four-Week Revision Data

Week	Feedback type	Holistic score (out of 6)		Times submitted		Word count		Time spent (in minutes)	
		S	*A*	*S*	*A*	*S*	*A*	*S*	*A*
Week 1	L2	6	5	6	6	363	296	90	63
Week 2	L1 & L2	5	5	7	7	334	361	77	72
Week 3	L2	6	6	5	4	318	343	76	61
Week 4	L1 & L2	6	6	6	6	327	365	100	63

Note. S = Stephen; A = Amos.

holistic scores for the last two weeks while their highest submissions were in week 2. Their word counts only varied by 23 words overall. While this category may seem trivial, Lee, Gentile, and Kantor (2010) found that a high word count often "co-occurs with other highly valued aspects of essay quality captured through holistic and multi-trait scoring rubrics" (p. 409), and Enright and Quinlin (2010) found that it is valued by human raters as well. Stephen worked on his essays a total of one hour and eleven minutes longer than Amos worked on his; likely accounting for this difference is that Stephen stayed after class each week to continue working, which was allowed since the study was classroom-based and not experimental.

Both students left only two errors uncorrected on English-only weeks in grammar, usage, and mechanics; Stephen began with 15 and Amos began with 30. On English + L1 weeks, Stephen left 3 uncorrected and Amos left 2, so their unattended errors over all weeks were virtually equal. In the repetition of words subcategory, both had very near 150 marks over the four weeks. Stephen reduced from 146 marks to 77, whereas Amos only reduced his 152 marks by 4.

Preferences Similar; Revision Focuses Differ

While their basic data from Criterion was very similar, their revision behaviors reveal several areas of interest: they displayed similar preferences for and uses of the L1, but the amounts of use differed: one focused on correction, and the other focused on revision; and, one displayed diligence, while the other displayed confidence.

Use of L1 Similar

While they both used the L1 glossed feedback, mainly in the Writer's Handbook and in bilingual dictionaries, Stephen used both of these more often. Even on weeks where he did not have access to the L1 glossed feedback,

Stephen opened a bilingual dictionary at the beginning of class, seemingly planning to use it. In week 2, he scrolled slowly through the Chinese portion of the explanation, moved very quickly through the English, and slowed down again when he came to Chinese (week 2, Camtasia). After using the L1 glossed feedback extensively, he described his corrections as "more" and "quicker" (week 2, interview). When given the choice, Stephen showed a pattern of using the L1 glossed feedback over the English-only feedback.

Amos's use of the L1 glossed feedback was similar, though less extensive. He especially used it in the Writer's Handbook, demonstrating the identical behavior that Stephen demonstrated, speeding through the English and slowing for the Chinese; he seemed to show a greater interest in the written feedback with the L1 glosses. Though both reported preferring the visual feedback, Amos checked the green box feedback with L1 glossed feedback over 20 times in the final week, though briefly. His use of the written feedback decreased in weeks without the L1 glossed feedback. In week 3, Amos gave a reason for this:

> STUDENT: If we have Chinese translation, use more Writer's Handbook.
> TEACHER: Ok, why?
> STUDENT: Because … some academic words in Writer's Handbook, we cannot understand English very easily, but in Chinese we can know what it say …
> TEACHER: Do you think it would encourage you to use the Handbook more if it had [the Chinese translations]?
> STUDENT: Yes.
>
> (week 3, interview)

Amos struggled, though, with his need for the L1, feeling he should not use it and looking to a time when he would no longer need it. In his final interview, he shared his belief that the L1 "is harmful for us to learn English," but he conceded that its "effect is some useful things. Just in the Writer's Handbook can give you more information for you to understand what they said" (week 4, interview). As this quote demonstrates, belief that the L1 should not be used or is harmful to learning is not only reserved for teachers, students may also feel this way (e.g., Liao, 2006).

Revision Focuses Differ

Stephen and Amos's revision times were used differently: while Stephen focused on correction, Amos focused on revision. To illustrate, one subcategory under "Style" marks overly repetitious words (Burstein & Wolska, 2003). Stephen lowered his repetition marks each week, once to zero. In week 2, after changing several words with help from a Chinese–English dictionary, Stephen sought advice from the teacher as to how many repetition errors are reasonable in an essay. It was suggested that he try to reduce his 30 marked repetitions

(301-word essay) by 10. He did refer to the Handbook, as suggested, reading the explanations slowly in Chinese and scrolling quickly through the English examples (week 2, Camtasia). This strategy seems to have worked for him because, at the end of class, Criterion did not identify any repetition errors in his final essay, even after his word count increased by 33 words. His work would suggest that he had intrinsic motivation to reduce these repetitions, as it was not required that he do so. Even when this category frustrated him in the third week, he was able to reduce his repetitions by 27%. On the Camtasia audio (week 3) he quietly laughed, "How to fix this problem?" Later in the interview, he was asked about this frustration:

> TEACHER: This repetition of words, was that where you got frustrated and you didn't know how to fix it?
> STUDENT: Yeah.
> TEACHER: And when you see so many blue highlights, how do you feel?
> STUDENT: Terrible. That mean a lot of mistake. Who want to make by his one essay? No one wants.
>
> (week 3, interview)

Stephen clearly describes how seeing that many marks may be disheartening for language learners, making his efforts in this category even more remarkable. Amos, on the other hand, appeared to expend no effort reducing repeated words; while Stephen took on this challenge, Amos ignored it. In the final week, 45 repeated words were marked, but he made no attempt to change any of these (week 4, Camtasia). The reason he gave for his neglect of this category was its difficulty (week 4, interview). Instead of focusing on correction, Amos spent time deleting, composing sentences, and adding details; namely, he spent his time revising.

Diligence versus Confidence Displayed

Stephen seemed to display unusual diligence in his work, looking for additional errors that Criterion had not marked. Many days, he stayed in the computer lab after class and worked on his essays; no other student in the class did this. These behaviors seem to show motivational intensity and positive self-efficacy. This is not to say Stephen is confident in his writing, but using the L1 feedback in AWE and outside sources seems to have bolstered his motivation. His persistence may demonstrate possible effects of utilizing the L1, having a better grasp of the task and better execution of it than he would have had alone (Storch & Wigglesworth, 2012).

Amos's behavior seemed to show a greater general confidence about his decision making; he often chose to correct errors differently than was suggested by Criterion. In one instance, Criterion gave him an incorrect error code, yet

he was able to correctly fix the true error. The original sentence was, "I am firmly convinced that every individual could learn a lot from they mistakes," and Criterion marked it as a missing comma error. Amos consulted the green box feedback that said, "You may need to place a comma after this word." He did not add a comma, rightly so, and then changed "they" to "their," successfully making, then, a correct fix in response to incorrect feedback. In week 3, Amos began by making corrections autonomously, even without the Chinese; for example, he immediately deleted "more" after seeing that the phrase "more better" was highlighted, without checking written feedback (week 3, Camtasia). He did use the written feedback later, checking a preposition error. After scrolling through the handbook and not finding what he was looking for, he made a successful change. Several other times, Amos checked the green box feedback and subsequently made correct revisions. However, on three occasions, he read the green box feedback and yet did not make any changes, though he should have (week 3, Camtasia). Whether successful or not in a given instance, being willing to go against Criterion's feedback seems to show a rather high level of confidence in his own English knowledge and decision-making ability.

Neither student ignored the L1 glossed feedback; both used the same features, sometimes in identical ways. However, Amos did seem to use the feedback less overall, perhaps preferring independence. Amos attempted more revision than Stephen did, often spending a majority of the time writing additional sentences rather than spending time on error correction, especially in the repetition-of-words category.

Concluding Remarks

Though it is small in scale, this type of study gives an important and necessary in-depth view of students in a classroom setting. Classroom case studies come with inherent limitations. The small scale may be a concern; however, since it was an actual classroom study, students could not be required to participate in the out-of-class interviews. Additionally, the two students who volunteered had similar backgrounds in English and were from the same country. Participants with varied levels and L1s could be desirable to compare differences in perceptions and performance. Since revisions were made in an intact classroom with non-controlled variables, individual revisions could not be definitively linked to the L1 feedback. Additional detail is available in the larger study (Wilken, 2013), if longer overall descriptions are desired.

Suggestions for Further Study

Following a progression of students or classes would be a way to track changes in perceptions over time (Bitchener, 2008; Guénette, 2007). Having participants of varied proficiency levels could also make a clearer delineation between the

levels who still feel they need the L1 glossed feedback and those who no longer feel the need for it. Another option would be to conduct an experimental study seeking to link revisions to the L1 feedback in a setting without the aid of teachers, peers, and other online helps, as were available in this classroom. In such a study, uptake rates of the L1 feedback could be determined for different error types as well.

AWE developers should continue to develop L1 glossed feedback in their programs. Further enhancements might be: (1) to offer a broader range of L1 feedback choices, (2) to refine the L1 translations, and (3) to provide a mechanism for the students to turn on and off the L1, giving them more control over their own learning.

This study asked two questions regarding the L1: first, what were the students' perceptions of the L1 feedback, and second, what did they do with this feedback? The students' feedback preferences were similar while the focus of their revisions differed. Stephen's focus was error reduction, whereas Amos's focus was content revision. Though the focus participants differed in their use and desire for the L1, both found it helpful, especially when using certain text-heavy features (i.e. the Writer's Handbook); Stephen, however, relied on it to a greater degree. When correcting errors, both preferred and used the visual cues of the blue highlights more than the written feedback found in the green box feedback and the Writer's Handbook. Finally, these students felt the L1 glossed feedback aided them in understanding and autonomy, frequently mentioning finding greater autonomy through AWE and the L1 glossed feedback.

As educators, we strive for ever-increasing clarity about how to provide CF effectively to our students. Since it is our task to provide learners with "easily accessible glosses" (Hulstijn, Hollander, & Greidanus, 1996, p. 337), we should consider including AWE and L1 glossed feedback in our CF arsenals. Folse (2004) believes a "brief translation of a key concept at the right time can be invaluable" (p. 60), and for this level of students, the L1 may expedite their revisions. Although there are many avenues to pursue to understand this issue fully, it seems that AWE, through its various forms of feedback, including L1 glossed feedback, gave these IEP students the feeling that they were moving toward the goal of becoming independent L2 writers. This increased sense of autonomy in using the target language should not be dismissed as being a non-necessity for ESL students; it seems, rather, that it could be a pivotal factor in determining future motivation and even success in their language learning.

About the Author

Jayme Wilken is a lecturer at Iowa State University, teaching in the Intensive English and Orientation Program and in the English Department. Her specialization is in computer assisted language learning, and her research focuses on using

technology in classroom settings. Jayme has presented at regional, national, and international conferences, including TESOL, CALICO, and WorldCALL.

References

Antón, M., & DiCamilla, F. (1998). Socio-cognitive functions of L1 collaborative interaction in the L2 classroom. *Canadian Modern Language Review/La Revue canadienne des langues vivantes, 54*(3), 314–342. https://doi.org/10.3138/cmlr.54.3.314

Atkinson, D. (1987). The mother tongue in the classroom: A neglected resource? *ELT Journal, 41*(4), 241–247. https://doi.org/10.1093/elt/41.4.241

Attali, Y. (2004). Exploring the feedback and revision features of Criterion. Paper presented at the National Council on Measurement in Education, San Diego, CA.

Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of Second Language Writing, 17*(2), 102–118. https://doi.org/10.1016/j.jslw.2007.11.004

Brooks, F. B., & Donato, R. (1994). Vygotskyan approaches to understanding foreign language learner discourse during communicative tasks. *Hispania, 77*(2), 262–274. https://doi.org/10.2307/344508

Burstein, J., & Wolska, M. (2003, April). Toward evaluation of writing style: Finding overly repetitive word use in student essays. In *Proceedings of the tenth conference on European chapter of the Association for Computational Linguistics,* 35–42. https://doi.org/10.3115/1067807.1067814

Chen, I C. F. E., & Cheng, W. Y. E. (2008). Beyond the design of automated writing evaluation: Pedagogical practices and perceived learning effectiveness in EFL writing classes. *Language Learning & Technology, 12*(2), 94–112.

Cook, V. (2001). Using the first language in the classroom. *Canadian Modern Language Review/La Revue canadienne des langues vivantes, 57*(3), 402–423. https://doi.org/10.3138/cmlr.57.3.402

Cotos, E. (2011). Potential of automated writing evaluation feedback. *CALICO Journal, 28*(2), 420–459. https://doi.org/10.11139/cj.28.2.420-459

Enright, M. K., & Quinlan, T. (2010). Completing human judgment of essays written by English language learners with e-rater scoring. *Language Testing, 27*(3), 317–334. https://doi.org/10.1177/0265532210363144

Esterberg, K. G. (2002). *Qualitative methods in social research*. Boston, MA: McGraw-Hill.

Folse, K. (2004). *Vocabulary myths: Applying second language research to classroom teaching.* Ann Arbor, MI: University of Michigan Press. https://doi.org/10.3998/mpub.23925

Goh, S. C. (2002). Managing effective knowledge transfer: An integrative framework and some practice implications. *Journal of Knowledge Management, 6*(1), 23–30. https://doi.org/10.1108/13673270210417664

Grimes, D., & Warschauer, M. (2010). Utility in a fallible tool: A multi-site case study of automated writing evaluation. *The Journal of Technology, Learning, and Assessment, 8*(6), 1–42. Retrieved May 2016 from http://www.jtla.org.

Guénette, D. (2007). Is feedback pedagogically correct?: Research design issues in studies of feedback on writing. *Journal of Second Language Writing*, *16*(1), 40–53. https://doi.org/10.1016/j.jslw.2007.01.001

Hegelheimer, V., & Tower, D. (2004). Using CALL in the classroom: Analyzing student interactions in an authentic classroom. *System, 32*(2), 185–205. https://doi.org/10.1016/j.system.2003.11.007

Hubbard, P. (2004). Learner training for effective use of CALL. In S. Fotos, and C. M. Browne (Eds.), *New perspectives on CALL for second language classrooms*. London: Routledge, 45–68.

Hulstijn, J. H., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *The Modern Language Journal, 80*(3), 327–339. https://doi.org/10.1111/j.1540-4781.1996.tb01614.x

Hyland, K., & Hyland, F. (2006). *Feedback in second language writing: Contexts and issues*. Cambridge, England: Cambridge University Press. https://doi.org/10.1017/CBO9781139524742

Jacobs, G. M., Dufon, P., & Fong, F. C. (1994). L1 and L2 vocabulary glosses in L2 reading passages: Their effectiveness for increasing comprehension and vocabulary knowledge. *Journal of Research in Reading, 17*(1), 19–28. https://doi.org/10.1111/j.1467-9817.1994.tb00049.x

Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. Beverly Hills, CA: Laredo Publishing Company.

Larsen-Freeman, D., & Long, M. (1991). *An introduction to second language acquisition research.* Applied linguistics and language study. London, England: Longman.

Laufer, B., & Shmueli, K. (1997). Memorizing new words: Does teaching have anything to do with it? *RELC journal, 28*(1), 89–108. https://doi.org/10.1177/003368829702800106

Lee, Y. W., Gentile, C., & Kantor, R. (2010). Toward automated multi-trait scoring of essays: Investigating links among holistic, analytic, and text feature scores. *Applied Linguistics, 31*(3), 391–417. https://doi.org/10.1093/applin/amp040

Levy, M., & Stockwell, G. (2006). *CALL dimensions: Options and issues in computer assisted language learning.* Mahwah, NJ: Lawrence Erlbaum Associates.

Li, J., Link, S., & Hegelheimer, V. (2015). Rethinking the role of automated writing evaluation (AWE) feedback in ESL writing instruction. *Journal of Second Language Writing, 27*, 1–18. https://doi.org/10.1016/j.jslw.2014.10.004

Liao, P. (2006). EFL learners' beliefs about and strategy use of translation in English learning. *RELC Journal, 37*(2), 191–215. https://doi.org/10.1177/0033688206067428

Link, S., Dursun, A., Karakaya, K., & Hegelheimer, V. (2014). Towards better ESL practices for implementing automated writing evaluation. *CALICO Journal, 31*(3), 323–344. https://doi.org/10.11139/cj.31.3.323-344

Mackey, A. & Gass, S. M. (2005). *Second language research: Methodology and design*. Mahwah, NJ: Lawrence Erlbaum Associates.

Prince, P. (1995). Second language vocabulary learning: The role of context versus translations as a function of proficiency. *The Modern Language Journal, 80*(4), 478–493. https://doi.org/10.1111/j.1540-4781.1996.tb05468.x

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129–158. https://doi.org/10.1093/applin/11.2.129

Storch, N., & Wigglesworth, G. (2010). Learners' processing, uptake, and retention of corrective feedback on writing: Case studies. *Studies in Second Language Acquisition, 32*, 303–334. https://doi.org/10.1017/S0272263109990532

Storch, N., & Wigglesworth, G. (2012). What role for collaboration in writing and writing feedback. *Journal of Second Language Writing. 21*(4), 364–374. https://doi.org/10.1016/j.jslw.2012.09.005

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics: Studies in honour of H. G. Widdowson* (pp. 125–144). Oxford: Oxford University Press.

Swain, M., & Lapkin, S. (2000). Task-based second language learning: The uses of the first language. *Language Teaching Research, 4*(3), 251–274. https://doi.org/10.1191/136216800125087

Truscott, J., & Hsu, A. Y. (2008). Error correction, revision, and learning. *Journal of Second Language Writing, 17*(4), 292–305. https://doi.org/10.1016/j.jslw.2008.05.003

Warschauer, M., & Ware, P. (2006). Automated writing evaluation: Defining the classroom research agenda. *Language Teaching Research, 10*(2), 157–180. https://doi.org/10.1191/1362168806lr190oa

Wilken, J. L. (2013). L1 feedback in automated writing evaluation: From learners' perspectives. Graduate Theses and Dissertations. Retrieved from ProQuest Digital Dissertations (1546461).

Yoshii, M. (2006). L1 and L2 glosses: Their effects on incidental vocabulary learning. *Language Learning and Technology, 10*(3), 85–101.