Meihua Liu
Tsinghua University

iD https://orcid.org/0000-0002-5951-4167

Focus and Effects of Peer and Machine Feedback on Chinese University EFL Learners' Revisions of English Argumentative Essays

Abstract

The present mixed-method study examined the focus and effects of peer and machine feedback on the revisions of English argumentative essays. The study collected data from 127 Chinese university EFL learners, which included Draft 1, peer feedback (PF), PF-based Draft 2, machine feedback (MF), MF-based Draft 2, questionnaires, and interview recordings. The main findings were: (a) peer feedback was primarily concerned with content errors while machine feedback mainly involved language errors, (b) significant differences occurred in most types of errors between Draft 1, PF and PF-based Draft 2, and between Draft 1, MF, and MF-based Draft 2, (c) the uptake of 'introducing a new topic in Conclusion' was a powerful predictor of PF-based Draft 2 scores, and (d) the participants generally moderately considered peer and machine feedback to be useful. Based on the findings, some implications are discussed on how to better implement and enhance the quality of peer and machine feedback.

Keywords: peer feedback, machine feedback, focus, effect, rewrite, argumentative essay

Introduction

As an essential component of students' academic development in a second/ foreign language (SL/FL), writing requires a considerable amount of time and effort since it involves higher order thinking, which makes it very challenging for many SL/FL writers (Cope et al., 2011; Dikli & Bleyle, 2014). Consequently, feedback plays a critical role in enhancing the quality of students' compositions. Nevertheless, assessing writing and providing feedback are also time-consuming and challenging. This is why though teacher feedback is more

effective (Goldstein, 2004; Hattie & Timperley, 2007; Keh, 1990; Sterna & Solomo, 2006; Vardi, 2009), machine and peer feedback has been developed and implemented in both classroom and other learning situations (Allen & Katayama, 2016; Shintani, 2015). Even though both peer review and machine feedback have proved to have positive effects on SL/FL learners' rewrites (Caulk, 1994; Hyland & Hyland, 2006; Rollinson, 2005; Rollinson, 1998, 2005; Topping, 1998; Yu & Lee, 2015), conflicts always exist about the actual effects (Anson, 2006; Xie, Ke & Sharma, 2008). Few studies have examined peer and machine feedback simultaneously either. Moreover, considering that accuracy is both an important and frustrating issue in writing (Li, Link & Hegelheimer, 2015), it is worthwhile to analyze more specifically the impact of peer and machine feedback on the quality of SL/FL learners' rewrites. For these reasons, the present mixed-method study, targeting Chinese university EFL (English as a FL) learners, explored the focus and effects of peer and machine feedback on learners' rewrites of English argumentative essays.

Literature Review

Defined as the "information with which a learner can confirm, add to, overwrite, tune, or restructure information in memory, whether that information is domain knowledge, meta-cognitive knowledge, beliefs about self and tasks, or cognitive tactics and strategies" (Winne & Butler, 1994, pp. 5740), feedback has been long held to facilitate the learning of SLs/FLs (Ellis, 2011; Ferris, 2010; Hattie & Timperley, 2007).

Peer Feedback

Student peer assessment (review/feedback) (PA) refers to an educational arrangement in which students assess the quality of their peers' work and provide feedback (Dochy, Segers, & Sluijsmans, 1999). There have been many studies on the results of PA in relation to grading and student satisfaction, as well as on effective organization of PA in higher education (Cho & Schunn, 2005; Gielen et al., 2010; Ion, Barrera-Cerominas & Tomàs-Folch, 2016; Kulkarni, Bernstein & Klemmer, 2015; Lin & Yang, 2011; Rollinson, 1998, 2005; Topping, 1998; Xie et al., 2008). Though some studies reveal that PA has little or even negative effect on SL/FL learners' writing (Xie et al., 2008), more studies show that peer readers can provide useful feedback (Caulk, 1994; Rollinson, 1998, 2005; Topping, 1998; Yu & Lee, 2015). Topping's (1998) review indicated that

PA was of adequate reliability and validity in a wide variety of applications and had positive formative effects on student achievement and attitudes. Ion et al.'s (2016) analyses of 637 feedback units showed that peer feedback helped students better develop the task in their writing.

In addition, trained PA can be more effective (Ellis, 2011; Kulkarni et al., 2015; Min, 2006). For example, Min (2006) examined the impact of trained responders' feedback on EFL college students' revisions in terms of revision types and quality. After a four-hour in-class demonstration of how to do peer review and a one-hour after-class reviewer-teacher conference with 18 students, the instructor-researcher collected students' first drafts and revisions, as well as reviewers' written feedback, and compared them with those produced prior to training. The results indicated that students incorporated a significantly higher number of reviewers' comments into revisions after the peer review training, and that the number of revisions with enhanced quality was significantly higher than that before the peer review training. The researcher thus concluded that trained peer review feedback could positively impact EFL students' revision types and quality of texts, supported by a subsequent study (Liu & Chai, 2009).

Moreover, peer feedback proves to be beneficial to students in other aspects (Ellis, 2011; Kurt & Atay, 2007; Lundstrom & Baker, 2009; Miao et al., 2006). Miao et al. (2006) examined peer and teacher feedback on essays of the same topic written by Chinese University EFL learners. Analyses of student texts, questionnaires, video recordings and interview transcripts revealed that peer feedback improved student autonomy thought it was less adopted in students' rewrites. Kurt and Atay's (2007) eight-week experimental study of 86 Turkish prospective teachers (PTs) of English showed that the peer feedback group experienced significantly less writing anxiety than the teacher feedback group at the end of the study. The study also revealed that the peer feedback process helped the PTs become aware of their mistakes and helped them look at their essays from a different perspective. Lundstrom and Baker (2009) did a study with 91 university students in nine writing classes at two proficiency levels to see which was more beneficial to improving student writing: giving or receiving peer feedback. The results indicated that the givers, who focused solely on reviewing peers' writing, made more significant gains in their own writing over the course of the semester than did the receivers, who focused solely on how to use peer feedback.

Machine Feedback

As technology develops, machine feedback becomes possible via computers and internet. The technology often used for feedback on writing is Automated Writing Evaluation (AWE) software which generates automated scores based

on techniques such as artificial intelligence, natural language processing and latent semantic analysis (Philips, 2007; Shermis & Burstein, 2003; Ullmann, 2019), and provides written feedback in the form of general comments, specific comments and/or corrections (Stevenson & Phakiti, 2014). In recent years, using AWE to provide feedback in the writing classroom has steadily increased, such as Project Essay GraderTM (PEG), e-rater, Intelligent Essay AssessorTM (IEA), and IntelliMetricTM (Stevenson & Phakiti, 2014). In China, the most widely used is www.pigai.org.[1] While many scholars applaud AWE as a means of freeing instructors from marking assignments and enabling them to devote more to writing instruction (Hyland & Hyland, 2006; Philips, 2007; Ullmann, 2019), others doubt whether AWE is capable of providing accurate and effective feedback (Anson, 2006).

For example, Li et al. (2015) used mixed-methods to investigate how Criterion1 affected writing instruction and performance. Four ESL writing instructors and 70 non-native English-speaking students participated in the study. The results showed that Criterion1 led to increased revisions and that the corrective feedback from Criterion1 improved accuracy from a rough to a final draft. AbuSeileek and Abualsha'r (2014) investigated the effect of computer-mediated corrective feedback on 64 EFL learners' performance in writing over the course of eight weeks. The participants were randomly assigned to either a no-feedback control condition or a corrective feedback condition. The researchers found that students who received computer-mediated corrective feedback while writing achieved better results in their overall test scores than students in the control condition who did not receive feedback. Cheng (2017) employed a mixed-method to investigate the impact of online automated feedback (OAF) on the quality of 138 university students' reflective journals in a 13-week EFL course. The findings showed that the experimental group outperformed the control group in the overall score of the final reflective journal and demonstrated a significant improvement in scores across reflective journals. The results of these two studies show that AWE has a positive impact on the quality of students' writing, supporting those of earlier studies (Chen & Cheng, 2008; Warschauer & Ware, 2006). Ullmann's (2019) study of 76 essays showed that the automated analysis was immediate, scalable, and only on average 10% less accurate than the manual analysis.

Even so, Stevenson and Phakiti's (2014) review found little evidence for positive effects of AWE on the quality students' rewrites based on AWE. Stevenson and Phakiti (2014) attributed this to little research, heterogeneity of existing research, the mixed nature of research findings, and methodological issues. Other explanations are that computers do not possess human

---

[1] The most popular online platform which provides feedback on and assesses English writing automatically in China. Users create accounts on it first, upload their writing and then receive feedback immediately on it.

inferencing skills and background knowledge (Anson, 2006) and that AWE-generated comments primarily focus on grammar in writing (Hyland & Hyland, 2006). This may be why AWE-generated feedback is less acceptable to students than teacher feedback (Dikli & Bleyle, 2014). Dikli and Bleyle (2014) investigated the use of an AES system on 14 advanced students from various linguistic backgrounds in a college ESL writing classroom. The findings showed that the instructor provided more and better quality feedback and the AES system.

Rationale for the Study

As reviewed, there have been many studies on the results of peer and machine feedback in relation to grading and students' compositions (Bijami et al., 2013; Cho & Schunn, 2005; Gielen et al., 2010; Kulkarni et al., 2015; Lin & Yang, 2011; Rollinson, 1998, 2005; Topping, 1998; Xie et al., 2008). However, little has been said as to the focus of peer and machine feedback in educational designs (AbuSeileek & Abualsha'r, 2014). Few studies have simultaneously examined peer and machine feedback either. More insight into the nature of peer and machine feedback would indicate more clearly how technology and students could be more helpful in SL/FL writing and what kind of assistance teachers should preferably provide. For example, if technology and peers can provide useful feedback on grammar, teachers can direct their assistance more to textual coherence or content (AbuSeileek & Abualsha'r, 2014). Moreover, since writing accuracy is both an important and frustrating issue (Li et al., 2015), it is worthwhile to examine more specifically the focus and effects of peer and machine feedback on the quality of SL/FL learners' writing. For these reasons as well as the intent to make better use of peer and machine feedback, the present study adopted mixed methods to explore the focus and effects of peer and machine feedback on Chinese university EFL learners' rewrites of English argumentative essays. To achieve this purpose, the following research questions were formulated:

(1) What is the respective focus of peer and machine feedback on students' English argumentative essays?

(2) How does peer and machine feedback impact students' rewrites of English argumentative essays?

Research Design

Context

The present research was conducted in a highly accredited university in Beijing, where English reading and writing courses were compulsory to undergraduate non-English majors. Upon entering the university, all non-English majors took a standardized English placement test, the results of which put the students into three band levels (a higher band level meant higher English proficiency). Based on their band levels, the students registered in compulsory and optional English courses accordingly. The majority fell into band level 2 and were required to take the English Argumentative Reading and Writing course, which contextualized the present study. The respondents of this study were randomly selected from those registered in the course taught by the same instructor. The students met the instructor once a week for a 90-minute period, who were required to write three long argumentative essays (more than 400 words) as well as a few short ones (about 100 words) during the 16-week semester. The instructor, PhD in Applied Linguistics, had been publishing widely in international journals and teaching the course for five years. In class, the students and the instructor discussed the techniques related to English argumentative essay reading and writing such as text structure, statement of arguments, paragraph structure, argument-developing skills, use of evidence, cohesion and coherence, and use of references. Adopting the process approach to writing, the instructor stressed the importance of revision and encouraged students to revise their drafts on the same composition at least twice from different sources: teacher feedback, peer comments and machine feedback. Prior to writing, a 30-minute peer review training based on Kramer, Leggett and Mead's scheme (1995) was arranged in class, which covered both content and language errors with more focus on content errors in that students had learned English grammar systematically but had not been trained how to write English argumentative essays effectively in previous schooling. Then students practiced peer review for each subsequent assigned writing task. Once a writing assignment was finished, each student sent his/her writing to the instructor, a peer, and www.pigai.org, independently. The instructor provided feedback electronically on each draft at sentence, paragraph and text levels, then gave a 25-minute summary report of the feedback and had individual discussions about the feedback when required by the students in the subsequent class; students assessed their peers' writing either electronically or in paper and must finish it within two days upon receiving the writing; www.pigai.org generated feedback in both Chinese and English (namely,

machine feedback in the present research) immediately upon receiving the submission. To avoid cross impact, students were required to revise their writing separately upon receiving different types of feedback.

Participants

127 (102 male and 25 female) students participated in the present study and answered the questionnaires related to their background information and perceptions of peer and machine feedback, of whom 64 were interviewed for their verbal perceptions about peer and machine feedback. Meanwhile, the first and second drafts of the same composition of 111 students, as well as peer and machine feedback, were complete for analyses. With an age range of 16–27 and an average of 19.42, the participants were from various disciplines such as civil engineering, mathematics, chemistry, and architecture. Prior to the course, they had never taken an English Argumentative Writing course.

Instruments

The collected data in the present study included interview transcripts, peer feedback (PF), machine feedback (MF), student draft 1, PF-based draft 2, MF-based draft 2, and writing scores, as detailed below.

Student texts. Draft 1, peer feedback, PF-based Draft 2, machine feedback, and MF-based Draft 2 of the course's second composition on global warming were collected. Based on student consent and the completeness of both drafts, 111 compositions of each draft as well as peer and machine feedback were finally collected for analyses.

Writing scores. The scores of each draft were collected, which was rated by the instructor on a scale of 1–15 in terms of text structure, power of argumentation, coherence, grammar and use of words (Appendix I).

Perceptions of peer and machine feedback questionnaire. This 14-item Perceptions of Peer and Machine Feedback Questionnaire (PPMFQ) was self-developed to investigate students' attitudes towards peer and machine feedback in terms of their roles and usefulness in their composition revisions. The questionnaire involved such issues as grammar, use of words, expression of viewpoints, use of evidence and references, which are crucial elements of argumentative essays (Wyrick, 2008). All the items were placed on a 7-point Likert Scale, ranging from 'Strongly Disagree' to 'Strongly Agree' with values of 1–7 assigned to each of the alternatives respectively.

Informal semi-structured interview. The informal semi-structured interview guide covered such questions concerning teacher feedback, peer and

machine feedback, their advantages, disadvantages and effects on composition revisions.

The background questionnaire. The background questionnaire aimed to collect informants' personal information such as age, gender, and major.

Procedure

Data were collected during weeks 7–9 of the semester when the second argumentative essay on global warming was assigned with the instructor's consent. To help students better understand the nature of argumentative essays, prompts on the task were provided such as effects of global warming on agriculture and major cause for global warming. Draft 1 was finished and submitted to the instructor, peers and www.pigai.org online (an account was created for the class beforehand) in week 7, followed by peer feedback within two days and immediate machine feedback, respectively. Based on the feedback, students revised their Drafts 1 independently according to the peer and machine feedback they had received respectively, and then submitted the rewrites to the instructor thereafter. Piloted to two students who had took the same course in the previous semester, the questionnaire was slightly modified, and then distributed to students together with a consent form who answered them in about 10 minutes in week 9's class meeting. According to their consent forms, a total of 64 students was informally interviewed by two research assistants thereafter in week 9. Each time, two students were interviewed together, which was mainly conducted in Chinese, recorded and lasted for 15–20 minutes.

Data Analyses

Since a writer needs to utilize an established language system to organize and present ideas in a certain mode in writing, the present study analyzed student texts and feedback in terms of both grammar and content. For this purpose, this study categorized errors with reference to the revision scheme in Kramer et al. (1995). The scheme (see Appendix II) used in the present study covered four types of errors: content errors (nine aspects involving failure to show a controlling idea, improper topic sentence and failure to achieve paragraph coherence, etc.), mechanical errors (misspelling, punctuation, and capitalization errors), syntactical errors (errors involving tense, part of speech, article, verb, adjective/adverb degree, agreement, and case, etc.), and lexical errors (errors in word formation, word choice, collocation, and unclear expression). Draft 1, PF-based Draft 2 and MF-based Draft 2 were

analyzed carefully according to the scheme to identify the errors students made in their writing. All the analyses were done by two research assistants with an overall inter-rater coefficient of .9l. Then the number of each type of error was counted for each text. The results were then analyzed via SPSS 20 to explore the distribution of and differences in different types of errors between Draft 1, peer feedback, PF-based Draft 2, machine feedback and MF-based Draft 2. To explore the effects of peer feedback on student revisions, Draft 1 and PF-based Draft 2 were compared to count and compute the uptake of peer feedback in the corresponding rewrites, so were Draft 1 and MF-based Draft 2. Then, multiple regression analyses were run, with scores of PF-based and MF-based Draft 2s being the dependent variable and the uptake of peer and machine feedback of errors of different types being independent variables.

The survey data were computed via SPSS 20. The mean and standard deviation of each survey item were computed to determine how students perceived peer and machine feedback respectively. The interview recordings were first transcribed, double-checked and then subjected to thematic content analyses by the two research assistants respectively with an inter-rater reliability of .932 (Charmaz, 2006). The themes were then generalized, counted, and supported with excerpts from the interviewees' comments. Example themes were strengths of peer feedback, weaknesses of machine feedback, benefits of peer and machine feedback. When reporting the comments, a number was used for each interviewee for the sake of privacy and convenience.

Results

Text Analyses Results

Distribution of errors. Preliminary analyses of peer feedback showed that students commented on content errors in specific places of their peers' writing but provided very general comments on language problems such as 'There are lots of grammatical errors in the essay' in the writing. By contrast, www.pigai.org generated fairly specific suggestions on language problems but offered no content-related suggestions in students' writing. Consequently, further analyses of PF and PF-based Draft 2 focused on content errors while those of MF and MF-based Draft 2 focused on language errors. The errors in Draft 1, PF, PF-based Draft 2, MF, and MF-based Draft 2, were coded and counted, which were then analyzed in terms of mean and standard deviation (see Table 1).

Table 1

Means and standard deviations of errors in student texts, PF and MF (N = 111)

Error types	Draft 1		PF		PF-based Draft		MF		MF-based Draft 2	
	M	SD	M	SD	M	SD	M	SD	M	SD
C1	.62	.68	.095	.30	.14	.35				
C2	.50	.76	.65	.72	.25	.51				
C3	1.19	1.43	.32	.47	.54	.53				
C4	.36	.74	.16	.41	.16	.41				
C5	.21	.45	.02	.13	.14	.35				
C6	.21	.47	.03	.18	.095	.30				
C7	.299	.46	.54	.69	.19	.396				
C8	.31	.46	.87	1.02	.21	.41				
C9	.25	.43	.00	.00	.02	.13				
TotalC	3.96	2.81	2.68	1.68	1.75	1.37				
ME	1.07	2.41					.91	2.49	.51	.84
SS1	1.19	1.76					.397	.64	.856	1.27
SS2	1.68	1.77					.286	.96	1.22	1.62
SS3	1.22	1.21					1.79	1.68	.48	.79
SS4	.83	1.18					.302	.59	.58	.91
SS5	.09	.34					.064	.25	.04	.198
SS6	2.67	2.22					.73	.95	2.20	2.22
SS7	1.49	1.51					.524	.95	.92	.95
SS8	.15	.42					.032	.25	.102	.31
SS9	.09	.29					.00	.00	.02	.14
SS10	.55	1.09					.032	.25	.469	.92
SS11	.79	.94					.206	.48	.469	.74
SS12	.11	.47					.00	.00	.00	.00
SS13	.26	.79					.048	.28	.265	.73
SS14	.07	.25					.064	.30	.082	.28
SS15	.14	.51					.032	.18	.327	.63
SS16	.73	.82					.095	.39	.56	.79
TotalSS	13.09	5.96					4.97	3.35	8.46	4.16
LE1	.06	.27					.00	.00	.02	.14
LE2	2.13	2.02					.238	.67	1.52	1.89
LE3	1.25	1.25					1.44	1.47	.92	1.12
LE4	1.25	1.11					.016	.13	.76	.87
TotalLE	4.68	2.32					1.698	1.58	3.22	2.09
grammar							.365	.87	.00	.00
TotalE	21.74	8.15					7.03	4.28	13.72	5.67
Writing score	11.38	1.83					11.77	1.67	11.48	1.789

Notes: Please refer to Appendix I for the abbreviations of error types.

TotalC = total number of content errors; TotalSS = total number of syntactic errors; TotalLE = total number of lexical errors; TotalE = total number of errors

As seen from Table 1, the errors with highest mean scores in Draft 1 were SE6 (article errors) (mean = 2.67), LE2 (word choice errors) (mean = 2.13), SE2 (tense errors) (mean = 1.68), SE7 (errors of plural or singular nouns) (mean = 1.49), LE3 (collocation errors) (mean = 1.25), LE4 (unclear expressions) (mean = 1.25), SE3 (agreement errors) (mean = 1.22), SE1 (errors in part of speech) (mean = 1.19), C3 (failure to provide adequate evidence) (mean = 1.19), and ME (mechanical errors) (mean = 1.07). Peer feedback predominantly focused on content errors, barely involving syntactic errors except for such comments as "there are many tense errors in the writing" or "grammatical errors are too many" (comments like these were not counted in the final analyses in the paper because they were not specific). The means of content errors ranged from 0 (C9-introducing a new topic in Conclusion) to 1.02 (C8-inconsistency between the conclusion and the main argument). On the other hand, machine feedback was solely concerned with mechanical, syntactic and lexical errors. The errors in MF ranged from 0 (SE12-illogical comparison or ill parallelism) to 1.79 (SE3), and errors with highest mean scores were SE3 (agreement errors) (mean = 1.79), LE3 (collocation errors) (mean = 1.44), ME (mechanical errors) (mean = .91), SE6 (article errors) (mean = .73), and SE7 (errors of plural or singular nouns) (mean = .524).

Since PF and MF focused on certain aspects of Draft 1, most of which were incorporated into respective rewrites, the analyses of Draft 2 focused on the type of feedback students received correspondingly. As reported in Table 1, the mean scores of content errors ranged from .02 (C9) to .54 (C3) in PF-based rewrites and from 0 (SE12) to 2.20 (SE6) in MF-based rewrites.

Comparison of mean scores of the errors across Draft 1, PF, and PF-based Draft 2 shows that all content errors scored the highest in Draft 1 and that most content errors scored higher in PF than in PF-based Draft 2. Paired samples t-test results (see Table 2) indicated that Draft 1 differed significantly from PF in all types of content errors except C2 (improper topic sentence/no controlling idea/no topic sentence), largely with a small or medium effect size. Namely, significantly more content errors of all types existed in Draft 1 than identified by peers. Table 2 also shows that PF differed significantly from PF-based Draft 2 in C2 (t = 3.97), C3 (failure to provide adequate evidence) (t = –2.50), C5 (lack of the power of the argument/weak arguments or evidence) (t = –2.65), C7 (fail to achieve paragraph coherence: poor organization/Lack or misuse of transitional markers) (t = 3.73), C8 (inconsistency between the conclusion and the main argument) (t = 4.66), and TotalC (t = 3.66). Alternatively, significantly more errors of C2, C7, C8, and TotalC (total content errors) were identified in PF than in PF-based Draft 2, but the latter had significantly more errors of C3 and C5 than in the former. Yet Draft 1 had significantly more errors of C1 (failure to show a controlling idea/More than one controlling idea) (t = 5.47), C2 (t = 3.16), C3 (t = 4.10), C7 (t = 2.3l), C9 (introducing a new topic in Conclusion) (t = 2.78), and TotalC (t = 5.88) than in PF-based Draft 2.

A similar pattern was observed for Draft 1, MF, and MF-based Draft 2, as reported in Table 1. Mechanical errors and most syntactic and lexical errors scored the highest in Draft 1, and errors of some types scored higher in MF than in MF-based Draft 2 while it was reversed for errors of other types. Paired samples t-test results (see Table 3) demonstrated that Draft 1 differed significantly from MF in all syntactic errors except SE5 (adjective/adverb degree errors), SE12 (errors of illogical comparison or ill parallelism), SE13 (errors of sentence fragments/run-on sentence/dangling modifiers), SE14 (errors of mixed or confused expression and sentence structure), SE15 (missing a part of the sentence), and all lexical errors except LE1 (errors in word formation) and LE3 (errors in collocations). Namely, significantly more errors of most types were identified in Draft 1 than in MF except SE3 (errors in agreement) and LE3. Table 3 also suggests that MF identified significantly more errors of SE3 but significantly fewer errors of SE1 (errors in part of speech), SE2 (tense errors), SE6 (articles errors), SE10 (errors in word order), SE11 (errors in coordinating conjunctions and subordinating conjunctions), SE16 (overuse of a part of the sentence), TotalSE (total syntactic errors), LE2 (errors in word choice), LE4 (unclear or incomplete expressions), TotalLE (total lexical errors), and TotalE (total errors) than in MF-based Draft 2. In addition, Draft 1 had significantly more errors in SE2 (tense errors), SE3 (errors in agreement), SE6 (articles errors), SE7 (errors in the use of plural or singular forms/uncountable nouns), SE11 (errors in coordinating conjunctions and subordinating conjunctions), SE15 (missing a part of the sentence), SE16 (overuse of a part of the sentence), TotalSE, LE2, LE3, LE4, TotalLE, and TotalE than in MF-based Draft 2.

Table 2

Paired samples t-test results of peer feedback (N = 111) (degree of freedom = 110)

	Draft 1 & PF			PF & PF-based Draft 2			Drafts 1 & PF-based 2		
	t	p	d	t	p	d	t	p	d
C1	5.17	.000	0.99	−.77	.443	/	5.47	.000	1.04
C2	−1.18	.241	/	3.97	.000	0.76	3.16	.002	0.60
C3	4.55	.000	0.87	−2.50	.015	-0.48	4.10	.000	0.78
C4	2.83	.006	0.54	.00	1.00	/	1.52	.133	/
C5	3.02	.004	0.58	−2.65	.010	-0.51	1.35	.182	/
C6	2.18	.033	0.42	.39	.698	/	.629	.531	/
C7	−2.795	.007	−0.53	3.73	.000	0.71	2.31	.024	
C8	−4.35	.000	−0.83	4.66	.000	0.89	1.43	.159	/
C9	3.00	.004	0.57	−1.00	.321	/	2.78	.007	0.44
TotalC	2.46	.017	0.47	3.66	.001	0.70	5.88	.000	1.12
Score							−1.52	.131	/

Notes: effect size of Cohen's d: small = d ≤ 0.2; medium = d = 0.5; large = d ≥ 0.8 (Cohen, 1988)

Table 3

Paired samples t-test results of machine feedback (N = 111) (degree of freedom = 110)

Error types	Draft 1 & MF			MF & MF-based Draft 2			Drafts 1 & MF-based 2		
	t	p	d	t	p	d	t	p	d
ME	.56	.58	/	1.01	.319	/	1.39	.172	/
SS1	3.10	.003	0.59	−2.40	.020	−0.46	1.37	.177	/
SS2	6.12	.000	1.17	−3.31	.002	−0.63	2.18	.034	0.42
SS3	−2.79	.007	−053.	5.13	.000	0.98	4.67	.000	0.89
SS4	2.25	.028	0.43	−1.70	.096	/	−.72	.472	/
SS5	.38	.709	/	1.43	.159	/	1.00	.322	/
SS6	6.20	.000	1.18	−3.99	.000	−0.76	2.99	.004	0.57
SS7	4.38	.000	0.84	−1.76	.084	/	2.31	.025	0.44
SS8	2.31	.024	0.44	−1.35	.182	/	.000	1.00	/
SS9	2.56	.013	0.49	−1.00	.322	/	1.77	.083	/
SS10	4.05	.000	0.77	−3.59	.001	−0.68	.28	.785	/
SS11	4.22	.000	0.80	−2.53	.015	−0.48	2.07	.044	0.40
SS12	1.84	.070	/	.00	.000	/	1.66	.103	/
SS13	1.90	.063	/	−1.75	.086	/	−.330	.743	/
SS14	.30	.766	/	−.33	.743	/	−.57	.569	/
SS15	1.52	.135	/	−3.10	.003	−0.59	−2.22	.031	−0.42
SS16	4.93	.000	0.94	−3.52	.001	−0.67	2.06	.044	0.39
TotalSS	9.79	.000	1.87	−4.36	.000	−0.83	5.81	.000	1.11
LE1	1.00	.321	/	−1.00	.322	/	−1.00	.322	/
LE2	6.65	.000	1.27	−4.15	.000	−0.79	3.21	.002	0.61
LE3	−1.13	.263	/	1.87	.068	/	2.78	.008	0.53
LE4	8.02	.000	1.53	−6.17	.000	−1.18	3.28	.002	0.63
TotalLE	9.35	.000	1.78	−4.14	.000	-0.79	4.91	.000	0.94
TotalE	14.32	.000	2.73	−6.02	.000	−1.15	7.83	.000	1.49
Score							−1.82	.091	/

Notes: effect size of Cohen's d: small = d ≤ 0.2; medium = d = 0.5; large = d ≥ 0.8 (Cohen, 1988)

Effects of peer and machine feedback on students' rewrites. To explore the effects of peer and machine feedback on students' rewrites, multiple regression analyses were run, with PF-based and MF-based Draft 2 scores being dependent variables and the uptake of errors of different types being independent variables respectively. Regression analyses yielded no model for MF-based Draft 2 scores and 1 model for PF-based Draft 2 scores, as shown in Table 4.

Table 4

Multiple regression coefficients and significance of error predictors for PF-based Draft 2 scores

Uptake of errors	β	*t*	p	df	VIF	Cohen's f²
C9	.261	2.11	.039	1	1.00	.012

Notes: df = degree of freedom effect size of Cohen's f2: small = f2 ≤ .02; medium = f2 = .15; large = f2 ≥ .35 (Cohen, 1988)

As shown in Table 4, with the change in R^2 being .068, C9 (introducing a new topic in Conclusion) was the only predictor (β = .261, t = 2.11, f^2 = .012) that positively predicted the scores of students' rewrites based on peer feedback.

Self-reported Results

Survey results. The mean and standard deviation of each survey item concerning peer and machine feedback were computed (see Table 5),

Table 5

Self-reported Questionnaire Result (N = 127)

Self-reported questionnaire items	PF		MF	
The feedback	Mean	SD	Mean	SD
1. improved my ability to use grammar correctly.	4.93	1.11	5.56	1.21
2. improved my ability to use vocabulary appropriately.	4.87	1.16	5.54	1.17
3. enhanced my knowledge of the structure of academic English argumentative essays.	5.28	1.15	4.06	1.46
4. improved my ability to state the main argument clearly in academic English argumentative essays.	5.22	1.18	3.95	1.43
5. improved my ability to state supporting arguments clearly in academic English argumentative essays.	5.24	1.07	3.94	1.50
6. enhanced the logic of arguing for points in my academic English argumentative essays.	5.26	1.03	3.94	1.46
7. improved the coherence and cohesion in my academic English argumentative essays.	5.01	1.14	4.38	1.47
8. improved my ability to cite properly in academic English argumentative essays.	4.67	1.17	4.06	1.58
9. improved my ability to use vocabulary formally in academic English argumentative essays.	4.63	1.23	5.08	1.19
10. improved my ability to argue adequately in academic English argumentative essays.	5.10	1.16	4.02	1.48
11. improved my ability to argue substantially in academic English argumentative essays.	5.29	1.12	3.97	1.55
12. improved my ability to use argument-developing skills in academic English argumentative writing.	4.97	1.27	4.25	1.45
13. was mostly incorporated into my revised draft.	5.33	1.18	5.20	1.35
14. was largely acceptable.	5.42	1.09	5.24	1.29

which shows that the students scored 4.63–5.42 on the Perceptions of Peer Feedback Questionnaire (PPFQ) items and 3.94–5.56 on the Perceptions of Machine Feedback Questionnaire (PMFQ) items. The five PPFQ items with the highest means were items 14 (acceptability of peer feedback) (mean = 5.42), 13 (uptake of peer feedback) (mean = 5.33), 11 (relevance between [main] claims and supporting evidence) (mean = 5.29), 6 (logic of arguing) (mean = 5.26) and 5 (statement of supporting arguments) (mean = 5.24), centering on content. The five PMFQ items with the highest means were items 1 (improved ability to use grammar) (mean = 5.56), 2 (improved ability to use vocabulary appropriately) (mean = 5.54), 14 (acceptability of machine feedback) (mean = 5.24), 13 (uptake of machine feedback) (mean = 5.33), and 9 (improved ability to use vocabulary formally) (mean = 5.08), centering on the use of expressions and grammar. These findings indicated that the students were generally moderately positive toward peer and machine feedback.

Interview results. Table 6 summarizes the interviewees' perceptions of the advantages and disadvantages of peer and machine feedback. As seen in Table 6, around 20% of the interviewees commented that peer feedback provided more communication (23.4%), more chances to learn from each other (21.3%), new perspectives (21.3%) and good advice on language use and sentence polishing (17%). According to the interviewees, peers "feel more at ease and communicate frequently when reviewing each other's writing. This helps us to understand each other's writing better" (No. 34), and could "identify problems in logic" (No. 22), peer review enabled "me to know others' views of my writing" (No. 46), and "me to be aware of similar mistakes in my own writing" (No. 51). Meanwhile, since "we peers are at a similar English proficiency level, most peer comments are not much professional or appropriate" (No. 53), and "it is difficult for us to offer specific suggestions" (No. 35).

As seen in Table 6, machine feedback could "identify language and grammar mistakes effectively" (No. 31), and "better the sentences and format in my writing" (No. 18). However, because it was a machine, it could not "identify logical problems" (No. 10) or offer any content-related suggestions on aspects like "paragraph structure, statements of main and supporting arguments, and use of evidence" (No. 25). Moreover, the machine frequently "misidentified mistakes" (No. 31).

Probably because of these reasons, 72.3% and 63.9% of the interviewees reported that peer and machine feedback was helpful to the revision of their writing, respectively. On the whole, 100% and 71.7% of the interviewees reported feeling satisfied with peer and machine feedback, respectively.

Table 6

Self-reported Perceptions of Peer and Machine Feedback (N = 64)

Feedback	Advantages	Disadvantages
PF	a) more communication (11/23.4%), b) chances to learn from each other (10/21.3%), c) new perspectives (10/21.3%), d) good advice on language use and sentence polishing (8/17%), e) suggestions being very specific (6/12.8%), f) being friendly (4/8.5%), g) feeling at ease (3/6.4%).	a) not being inclusive (15/31.9%), b) comments being not deep (12/25.5%), c) comments being not professional or appropriate (8/17%), d) time-consuming (4/8.5%).
MF	a) good and specific comments on vocabulary and grammar (38/80.9%), b) being timely (15/31.9%), c) being very convenient (6/12.8%), d) being very clear (5/10.6%).	a) having no content- related comments (40/85.1%), b) having a high mis-identification rate (13/27.7%).

Discussion

Focus of Peer and Machine Feedback

Analyses of the data showed that peer feedback primarily focused on content errors in the present study. Although the interviewees were intermediate-advanced learners, they were not confident enough to pinpoint language problems for their peers. This was also evident in the number of content errors they identified in PF, which was significantly lower than that in Draft 1. Apart from that, this might be partly attributed to the time-consuming nature of reviewing a text, which made the participants unwilling to provide detailed and specific suggestions. Meanwhile, as discussed in Yu and Lee (2015), EFL students' group peer feedback activities are often driven and defined by their motives, which are shaped and mediated by the sociocultural context. The learning context where the instructor emphasized content more than linguistic forms of argumentative writing might be partially accountable for the participants' performance in their PF in the present study. The students thus focused more on content errors correspondingly, which, nevertheless, needs to be further explored.

The present study also revealed that machine feedback was predominantly concerned with language errors, as found in Hyland and Hyland (2006). This might be because the so-called machine, though modeled on human intelligence, could still not detect human thinking to provide useful comments on contents of an essay. In addition, though it offered timely and generally accurate feedback

on language problems, it mistook the correct use of grammar and expressions to be incorrect or provided wrong suggestions for "correctly pinpointed mistakes" "at a rather high rate" (No. 62). For example, www.pigai.org marked the part 'will in' in the sentence "It will in turn lead to the large scale release of the greenhouse gas into the atmosphere" (Writing 44, Draft 1) to be wrong. This finding partially supports the view that AWE is incapable of providing accurate feedback in certain aspects (Anson, 2006). Hence, it is necessary for both instructors and learners to be cautious when utilizing machine feedback. This is especially so for learners with lower proficiency in the SL/FL who are more unlikely to distinguish wrongly identified errors by machines. Moreover, to what extent and what type of language use is identified as errors by machines need to be further researched.

Effects of Peer and Machine Feedback

Regressional analyses indicated that the uptake of 'introducing a new topic in Conclusion' was a significant predictor for students' PF-based rewrites. This might be related to the culture of writing in Chinese, which tends to bring about something new in concluding parts of an essay. This thus deserves attention in formal classroom teaching and the effects need to be further researched as well. Analyses of self-reported data showed that the participants were generally positive about peer feedback, as found in the current literature (Liu & Chai, 2009; Miao et al., 2006). Apart from positively affecting students' rewrites, peer feedback offered students chances to communicate with and learn from each other, to become (more) aware of their own mistakes, to look at their own writing from a new perspective, as found in some existing studies (Miao et al., 2006; Wang, 2014). Miao et al.'s (2006) study indicated that peer feedback helped promote student autonomy, especially in cultures which look up teachers as authority figures.

Self-reported data indicated that the participants were generally moderately positive towards machine feedback, commenting that it was good, specific, timely, clear and convenient. This suggests that machine feedback did have positive effects on the polishing of sentences in students' rewrites, consistent with the finding in many existent studies (Cheng, 2017; Hyland & Hyland, 2006; Li et al., 2015; Philips, 2007). On the other hand, machine feedback was sometimes wrong, which frustrated the participants in the present research. Because of this, students are advised not to solely rely on machine feedback and consult peers and/or the instructor when being unsure of the comments. These findings suggest that developers of such platforms/softwares have to enhance their reliability and validity and pay more attention to providing content-related feedback, which is of central importance to an essay. They also

indicate that EFL learners, especially low or low-intermediate learners, have to be cautious when using machine feedback. Writing instructors had better remind their students of this limitation of machine feedback. Otherwise, some feedback would be misleading and the uptake of such feedback would lead to (even worse) mistakes.

As illustrated in the present research, peer and machine feedback had positive effects on students' rewrites, at the same time they were not satisfactory in certain aspects. For example, peer feedback sometimes is not professional or appropriate, and superficial, as found in the present study. Thus, it is important to improve the quality of peer and machine feedback. As found in Yu and Lee (2015), student motives could have direct influence on students' participation in group peer feedback activities and their subsequent revisions. It is necessary to foster positive and constructive motives towards peer and machine feedback in students prior to revising the first drafts. Meanwhile, if peer feedback can be done anonymously, students may feel more comfortable in providing more and better feedback on different aspects of their peers' writing, as found in Lu and Bol (2007). If students become more proficient in the target language, they will be able to provide better feedback as well, so are they trained to provide peer feedback and to write (more) effectively. Integrating technology into the peer review process may also be beneficial to providing better and timely feedback (Ellis, 2011; Lin et al., 2011; Nobles & Paganucci, (2015). Nobles and Paganucci's (2015) mixed-method study of 18 high school students in a hybrid freshman English class at an independent school revealed that students perceived their writing to be of higher quality when writing with digital tools and that writing in online environments enhanced writing skill development. Kulkarni et al.'s (2015) study showed that students' final grades improved when feedback was delivered quickly, but not if delayed by 24 hours. In addition, it is equally important to train students to do peer review (Gielen et al., 2010; Liu & Carless, 2006; Rollinson, 1998). It is better for writing instructors to familiarize students with the peer review criterion and their expectations. As put in Stanley (1992, p. 230), "it is not fair to expect that students will be able to perform these demanding tasks [peer feedback] without first having been organized practice with and discussion of the skills involved." Strategies such as engaging students with criteria and embedding peer involvement within normal course processes may help promote peer feedback (Liu & Carless, 2006). Lastly, as found in Wang's (2014) investigation of 53 Chinese EFL learners' perceptions of peer feedback on their EFL writing over time, various factors affect students' perceived usefulness of peer feedback such as their knowledge of assigned essay topics, proficiency in the target language, attitudes, time constraints, and classroom environment. It is necessary for writing instructors to consider these factors when implementing peer feedback.

Conclusions

The present mixed-method study examined the focus and effects of peer and machine feedback on the rewrites of Chinese university EFL learners' English argumentative essays. The main findings were:

(1) peer feedback was primarily concerned with content errors, while machine feedback mainly involved language errors,

(2) significant differences occurred in errors of most types between Draft 1, PF and PF-based Draft 2, and between Draft 1, MF, and MF-based Draft 2,

(3) the uptake of 'introducing a new idea in Conclusion' was a powerful predictor of PF-based Draft 2 scores, and

(4) the participants generally moderately considered peer and machine feedback to be useful.

Although the present study yielded insightful findings, given that the participants were intermediate-advanced learners and the instructor was experienced in academic English writing, it is worth doing further research on different types of SL/FL learners and instructors to explore more about the focus and effects of peer and machine feedback. For example, lower proficient SL/FL learners may not be able to identify all language problems and/or distinguish correctly and incorrectly identified errors by machine; SL/FL learners with no/little training in argumentative writing may not be able to identify content errors. All these may not only lower the quality of peer feedback but also mislead learners to blindly depend on peer and machine feedback. More research on these issues with different SL/FL learner populations helps both learners and instructors to have a better understanding of peer and machine feedback. Then accordingly, peer and machine feedback may be better implemented to complement teacher feedback to improve the quality of SL/FL learners' writing as well as to alleviate writing teachers' workload.

Conflict of interest statement

On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

AbuSeileek, A., & Abualsha'r, A. (2014). Using peer computer-mediated corrective feed-back to support EFL learners' writing. *Language Learning & Technology, 18*(1), 76–95. http://llt.msu.edu/issues/february2014/abuseileekabualshar.pdf

Allen, D., & Katayama, A. (2016). Relative second language proficiency and the giving and receiving of written peer feedback. *System, 56*(1), 96–106.

Anson, C. M. (2006). Can't touch this: Reflections on the servitude of computers as readers. Machine scoring of student essays. In P. Freitag Ericsson & R. Haswell (Eds.), *Machine scoring of student essays* (pp. 38–56). Logan, Utah: Utah State University Press.

Bijami, M., Kashef, S. H., & Nejad, M. S. (2013). Peer feedback in learning English writing: Advantages and disadvantages. *Journal of Studies in Education, 3*(4), 91–97. https://doi.org/10.5296/jse.v3i4.4314

Caulk, N. (1994). Comparing teacher and student responses to written work. *TESOL Quarterly, 28*(1), 181–188.

Charmaz, K. (2006). *Constructing grounded theory*. London: Sage.

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

Cheng, G. (2017). The impact of online automated feedback on students' reflective journal writing in an EFL course. *The Internet and Higher Education*, *34*, 18–27.

Cho, K., & Schunn, C. D. (2005). Scaffolded writing and rewriting in the discipline: A web-based reciprocal peer review system. *Computers & Education, 48*, 409–426. https://doi.org/10.1016/j.compedu.2005.02.004

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cope, B., Kalantzis, M., McCarthey, S., Vojak, C., & Kline, S. (2011). Technology-mediated writing assessments: Principles and processes. *Computers and Composition, 28*, 79–96.

Dikli, S., & Bleyle, S. (2014). Automated Essay Scoring feedback for second language writers: How does it compare to instructor feedback? *Assessing Writing, 22*, 1–17.

Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: a review. *Studies in Higher Education, 24*, 331–350.

Ellis, M. J. (2011). Peer feedback on writing: Is on-line actually better than on-paper? *Journal of Academic Language & Learning, 5*(1), A88–A99.

Ferris, D. (2010). Second language writing research and written corrective feedback in SLA: Intersections and practical applications. *Studies in Second Language Acquisition, 32*, 191–201.

Gielen, S., Peeters, E., Dochy, F., Onghena, P., & Struyven, K. (2010). Improving the effectiveness of peer feedback for learning. *Learning and Instruction, 20*, 304–315. https://doi.org/10.1016/j.learninstruc.2009.08.007

Goldstein, L. M. (2004). Questions and answers about teacher written commentary and student revision: Teachers and students working together. *Journal of Second Language Writing, 13*(1), 63–80. http://dx.doi.org/10.1016/j.jslw.2004.04.006

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. https://doi.org/10.3102/003465430298487

Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. *Language Teaching, 39*, 83–101.

Ion, G., Barrera-Corominas, A., & Tomàs-Folch, M. (2016). Written peer-feedback to enhance students' current and future learning. *International Journal of Educational Technology in High Education, 13*(15). https://doi.org/10.1186/s41239-016-0017-y

Keh, C. L. (1990). Feedback in the writing process: a model and methods for implementation. *ELT Journal, 44*(4), 294–304.

Kramer, M. G., Leggett, G., & Mead, D. (1995). *Prentice hall handbook for writers.* Englewood Cliffs, NJ: Prentice Hall.

Kulkarni, C., Bernstein, M. S., & Klemmer, S. (2015). Peer studio: Rapid peer feedback emphasizes revision and improves performance. *Behavior, 2*(1), 14–18.

Kurt, G., & Atay, D. (2007). The effects of peer feedback on the writing anxiety of prospective Turkish teachers of EFL. *Journal of Theory and Practice in Education, 3*(1), 12–23. http://eku.comu.edu.tr/index/3/1/gkurt_datay.pdf

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.

Lin, W.-H., & Yang, S. C. (2011). Exploring students' perceptions of integrating Wiki technology and peer feedback into English writing courses. *English Teaching: Practice and Critique, 10*(2), 88–103. http://education.waikato.ac.nz/research/files/etpc/files/2011v10n2dial1.pdf

Liu, M., & Chai, Y. (2009). Attitudes towards peer review and reaction to peer feedback in Chinese EFL writing classrooms. *TESL Reporter, 40*(1), 33–50.

Liu, N. F., & Carless, D. (2006). Peer feedback: The learning element of peer assessment. *Teaching in Higher Education, 11*(3), 279–290.

Lu, R., & Bol, L. (2007). A Comparison of anonymous versus identifiable e-Peer review on college student writing performance and the extent of critical feedback. *Journal of Interactive Online Learning, 6*(2), 100–115.

Lundstrom, K., & Baker, W. (2009). To give is better than to receive: The benefits of peer review to the reviewer's own writing. *Journal of Second Language Writing, 18*, 30–43. https://doi.org/10.1016/j.jslw.2008.06.002.

Miao, Y., Badger, R., & Zhen, Y. (2006). A comparative study of peer and teacher feedback in a Chinese EFL writing class. *Journal of Second Language Writing, 15*, 179–200.

Min, H. T. (2006). The effects of trained peer review on EFL students' revision types and writing quality. *Journal of Second Language Writing, 15*, 118–141. https://doi.org/10.1016/j.jslw.2006.01.003.

Nobles, S., & Paganucci, L. (2015). Do digital writing tools deliver? Student perceptions of writing quality using digital tools and online writing environments. *Computers and Composition, 38*, 16–31.

Philips, S. M. (2007). *Automated essay scoring: A literature review* (SAEE research series #30). Kelowna, BC: Society for the Advancement of Excellence Education.

Rollinson, P. **(**1998). Peer response and revision in an ESL writing group: A case study. Universidad Autónoma de Madrid, Unpublished PhD thesis.

Rollinson, P. (2005). Using peer feedback in the ESL writing class. *ELT Journal, 59*(1), 23–30. https://doi.org/10.1093/elt/cci003.

Shermis, M. D., & Burstein, J. (Eds.). (2003). *Automated essay scoring: A cross-disciplinary perspective.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Shintani, N. (2015). The effects of computer-mediated synchronous and asynchronous direct corrective feedback on writing: A case study. *Computer Assisted Language Learning.* https://doi.org/10.1080/09588221.2014.993400.

Stanley, J. (1992). Coaching student writers to be effective peer evaluators. *Journal of Second Language Writing, 1*, 217–233.

Sterna, L. A., & Solomo, A. (2006). Effective faculty feedback: The road less traveled. *Assessing Writing, 11*, 22–41.

Stevenson, M., & Phakiti, A. (2014). The effects of computer-generated feedback on the quality of writing. *Assessing Writing, 19*, 51–65.

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research, 68*(3), 249–276. http://www.jstor.org/stable/1170598

Ullmann, T. D. (2019). Automated analysis of reflection in writing: Validating machine learning approaches. *International Journal of Artificial Intelligence in Education, 29*, 217–257. https://doi.org/10.1007/s40593-019-00174-2

Vardi, I. (2009). The relationship between feedback and change in tertiary student writing in the disciplines. *International Journal of Teaching and Learning in Higher Education, 20*(3), 350–361. http://www.isetl.org/ijtlhe

Wang, W. (2014). Students' perceptions of rubric-referenced peer feedback on EFL writing: A longitudinal inquiry. *Assessing Writing, 19*, 80–96. http://dx.doi.org/10.1016/j.asw.2013.11.008

Warschauer, M., & Ware, J. (2006). Automated writing evaluation: Defining the classroom research agenda. *Language Teaching Research, 10*(2), 1–24.

Winne, P. H., & Butler, D. L. (1994). Student cognition in learning from teaching. In T. Husen & T. Postlewaite (Eds.), *International encyclopedia of education* (2nd ed.) (pp. 5738–5745). Oxford, UK: Pergamon.

Wyrick, J. (2008). *Steps to writing well*. Beijing: Peking University Press.

Xie, Y., Ke, F., & Sharma, P. (2008). The effect of peer feedback for blogging on college students' reflective learning processes. *Internet and Higher Education, 11*, 18–25. https://doi.org/10.1016/j.iheduc.2007.11.001.

Yu, S. & Lee, I. (2015). Understanding EFL students' participation in group peer feedback of L2 writing: A case study from an activity theory perspective. *Language Teaching Research, 19*(5), 572–593.

Appendix I

Marking Criterion

12–15	8–11	4–7
• Clearly state the main idea of the paragraph, • Adequate evidence is presented, • Clear organization, • Coherently and grammatically presented	• State the main idea of the paragraph, • Some evidence is presented, • Good organization, • Adequately presented (i.e., not that coherently and grammatically)	• Vaguely state the main idea of the paragraph, • Little evidence is presented, • Poor organization, • Poorly presented (i.e., not coherently and grammatically)

Error Coding & Classification Scheme

Content errors	
C1	Failure to show a controlling idea/More than one controlling idea
C2	Improper topic sentence/No controlling idea/no topic sentence
C3	Failure to provide adequate evidence
C4	Failure to provide substantial evidence
C5	Lack of the power of the argument/Weak arguments or evidence
C6	Failure to keep the necessary consistency in meaning/Inconsistency between the topic sentence and supporting sentences
C7	Fail to achieve paragraph coherence: poor organization/Lack or misuse of transitional markers
C8	Inconsistency between the conclusion and the main argument
C9	Introducing a new topic in Conclusion
Mechanical errors (ME)	
ME1	Misspellings
ME2	Punctuation errors
ME3	Capitalization errors
Syntactical errors (SE)	
SE1	Errors in part of speech (noun/adj./adv./prep./pron./conj./verb)
SE2	Tense errors
SE3	Errors in agreement
SE4	Verb errors
SE5	Adjective/adverb degree errors
SE6	Articles errors
SE7	Errors in the use of plural or singular forms/uncountable nouns
SE8	Case errors
SE9	Errors in mood /auxiliaries (including modal auxiliaries)
SE10	Errors in word order (positive and negative sentence/questions/subordinate clause/ adverbs and adjectives)
SE11	Errors in coordinating conjunctions and subordinating conjunctions
SE12	Errors of illogical comparison or ill parallelism
SE13	Errors of sentence fragments/run-on sentence/dangling modifiers
SE14	Errors of mixed or confused expression and sentence structure
SE15	Missing a part of the sentence
SE16	Overuse of a part of the sentence
Lexical errors (LE)	
LE1	Errors in word formation
LE2	Errors in word choice
LE3	Errors in collocations
LE4	Unclear or incomplete expressions

Meihua Liu

Ziele und Ergebnisse des kollegialen und automatisierten Feedbacks bei der Korrektur englischsprachiger Argumentationsaufsätze von EFL-Studierenden einer chinesischen Universität

Zusammenfassung

In Anlehnung an eine empirische Untersuchung präsentiert der vorliegende Artikel die Ziele und Ergebnisse eines kollegialen und automatisierten Feedbacks (engl. *peer feedback*, *machine feedback*) bei der Korrektur des Inhalts eines englischsprachigen Argumentationsaufsatzes. An der Untersuchung nahmen 127 Studierende einer chinesischen Universität teil, für die Englisch eine Fremdsprache ist. Die gesammelten Daten entstammen verschiedenen Versionen des Aufsatzes (Arbeitsversion Nr. 1, kollegiales Feedback (PF), Arbeitsversion Nr. 2 – basiert auf kollegialem Feedback, automatisiertes Feedback (MF), Arbeitsversion Nr. 2 – basiert auf automatisiertem Feedback) sowie Fragebögen und aufgenommenen Interviews. Aus der Analyse erhobener Daten ergeben sich wesentliche Unterschiede in Bezug auf die erhaltenen Feedbacks, die im Text auf Grundlage verschiedener Versionen des Aufsatzes im Einzelnen erläutert werden. Es wird beispielsweise unter Beweis gestellt, dass das kollegiale Feedback in erster Linie auf inhaltliche Fehler konzentriert war, während sich das automatisierte Feedback meistens auf sprachliche Fehler bezog. In Anlehnung an die durchgeführte Analyse werden Implikationen betreffend kollegiales und automatisiertes Feedback präsentiert.

Schlüsselwörter: kollegiales Feedback, automatisiertes Feedback, Argumentationsaufsatz