Integrating Machine Translation Into EFL Writing Instruction: Process, Product and Perception

Jian Wang
School of Humanities, Geely University of China, Chengdu, China

Xinli Ke
Department of Foreign Languages, Southwest Jiaotong University Hope College, Chengdu, China

Abstract—**Although there is a great demand for machine translation (MT) among language learners, its potentials as a computer-assisted language learning aid remain under-explored. Against this backdrop, this study adopted a mixed research method and conducted a semester-long empirical investigation into how EFL learners in mainland China used MT to assist their writing, whether MT helped improve their writing competence and how they perceived MT in EFL writing instruction. The major findings comprise: 1) By using MT students made more lexical and grammatical changes in essay revision; 2) MT helped improve the learners' overall writing competence, and particularly had a greater effect on writing accuracy and lexical complexity than on other dimensions; 3) Students generally held a positive attitude towards incorporating MT into EFL writing instruction.**

Index Terms—**EFL writing instruction, machine translation, writing processes, written products, students' perceptions**

I. INTRODUCTION

Traditional EFL (English as a foreign language) writing instruction attaches greater importance to students' written products than to their writing processes or to their stances of a certain teaching approach. This pedagogical problem has been particularly prominent in mainland China's EFL teaching setting where English writing instruction "operates" mechanically by following the "task assignment--task completion--final draft submission" mode with delayed or even no corrective feedback from teachers or other sources. This long-standing product-oriented teaching tradition has its deficiencies and drawbacks, since it remains unknown how students employ meta-cognitive and cognitive writing strategies in the whole process. To remedy this situation, EFL writing instruction in the information age is required to cast away obsolete teaching methods and embrace computer-assisted language learning (CALL) tools as revolutionizing catalysts. What language teachers are faced with is how best to apply these tools to produce favourable outcomes for language learners. Among them, machine translation (MT) has gained popularity both inside and outside the classrooms.

The concept of MT was formally put forward by Warren Weaver in 1949 (Chéragui, 2012), which refers to the process of transforming a source language into a target language by using computer software generally compatible with the systems of personal computers and smart phones (Alhaisoni & Alhasysony, 2017). Presently, MT is seldom used in writing classrooms, largely due to its questionable reliability, as was pointed out by Bahri and Mahadi (2016) that MT engines might misuse lexical and grammatical items and couldn't infallibly convey cultural meanings. However, the tremendous demand for MT in the language learning context tells us quite another story, as language learners increasingly regard MT as a useful CALL aid for acquiring vocabulary, completing translation exercises, and finishing reading comprehension and writing tasks (Alhaisoni & Alhasysony, 2017). Specifically, MT has the following upsides: 1) aiding language learners in making preliminary translation and reducing cognitive load in the reading process (Baraniello et al., 2016; O'Brien et al., 2018); 2) providing learners with a target language model and deepening L2 writers' interlanguage knowledge in the lexical, semantic, syntactic and pragmatic aspects (Amaral & Meurers, 2011; Bahri & Mahadi, 2016; Bernardini, 2016); 3) helping students convey and communicate messages more smoothly in the process of L2 writing, improve their writing fluency, reduce errors in their essays, and provide "scaffolding support" in their writing process (Godwin-Jones, 2015); 4) offering direct feedback like lexical and syntactic alternatives, instead of just giving negative or general feedback (Lee, 2020); and 5) providing learners with a less threatening language learning environment, reducing their language learning anxiety, and enhancing their learning motivation and self-confidence (Kliffer, 2005; Niño, 2009). Undeniably, dramatic technological breakthroughs have enabled MT engines to provide more accurate and readable translations, so the effectiveness of MT cannot be completely ignored and proper use of MT is beneficial to language learners' learning processes.

To this end, this study conducts a preliminary investigation into MT's potentials as a CALL tool for ameliorating EFL learners' writing competence, learners' way of utilizing MT to facilitate the writing process, and their stances of introducing MT into writing instruction. Another rationale for this study consists in the scant combination of MT and writing instruction in mainland China' EFL teaching setting.

II. LITERATURE REVIEW

To learn about the state of the art in applying MT to L2 writing, we performed a retrieval on Web of Science and China National Knowledge Internet with the keywords of "MT+L2 writing" and "机器翻译+二语写作" respectively, only to find a handful of empirical studies having set foot on this territory. The following is a panorama of relevant literature directly associated with the current study.

A. MT in the L2 Writing Process

A usual way to comprehend students' writing process is to compare the first and final drafts of an essay and identify any changes made by the writers. Closely related to our area of interest is Lee's (2020) study which found that Korean EFL learners compared their manual and MT versions, noticed errors, identified alternative items, and rewrote some parts of their essays, and these processes assisted these learners in acquisition of grammatical knowledge, use of new words in appropriate contexts and retention of the newly learned items by using them in the final version. The results of Lee (2020) confirmed Wong and Lee's (2016) observation that language learning with MT could foster students' language awareness and noticing skills such as perceiving their lack of L2 knowledge and discerning corresponding items to be learned from the MT version which will be employed in new contexts to craft novel sentences. As Carroll and Swain (1993) claimed, error detection and correction improves grammatical accuracy in L2 writing and facilitates interlanguage development. Likewise, exposure to alternative language items can raise students' awareness that there are no one-to-one equivalents between the source and target languages (Baraniello et al., 2016).

Lee's (2020) study showcased the changes in students' first and final drafts and stated that the subjects commenced to center more on post-translation editing and began to shift their view of writing from a product to a process. The study further pointed out that the subjects even started to place more emphasis on their L1 writing after recognizing that their source text could determine the quality of the MT version. Despite such findings, this study did not display the writing processes in a real sense, or at least we have no way of knowing how the subjects were engaged in the writing activities with MT engines. Therefore, more in-depth information on research design should be straightened out.

B. Impacts of MT on Students' Writing Proficiency

Research shows that MT can promote writing, but most of the conclusions are merely based on open or semi-structured questionnaires. Very few studies have designed teaching experiments to explore MT's impact on learners' overall writing level. O'Brien et al. (2018) took 10 non-native English speakers from different research fields as the research subjects, and the research comprised two stages, lasting for six weeks. In the first stage, the subjects completed a 500-word abstract which was divided into two parts with similar length (one in subjects' mother tongues and the other in English). In the second stage, researchers used Google Translate to translate the mother tongue parts into English, then returned the complete English abstract to the subjects and asked them to revise the whole abstract. The results showed that MT and post-translation editing practices had no negative effects on the quality of academic writing.

Garcia and Pena (2011) required nine native English-speaking Spanish beginners to write in English first, translate English into Spanish with MT, and then edit the translated text, while the other seven native English-speaking Spanish intermediate learners wrote in Spanish directly. It was found that the scores of the L2 texts produced with the help of MT were higher than those produced directly in Spanish. The study also found that compared with intermediate learners who wrote directly in Spanish, junior learners writing with MT had fewer pauses in the writing process, indicating that MT could reduce L2 writers' cognitive load.

O'Neill (2012) divided 32 native English-speaking college French learners into two experimental groups and a control group. The first experimental group received training in MT and used translation software, the second experimental group was permitted to use MT but did not receive training, and the control group was neither trained nor allowed to use MT. All three groups of subjects completed two short essays, and the researchers reviewed the essays from six dimensions: comprehensibility, content, spelling, syntax, grammar and vocabulary. It was found that the total scores of the first experimental group in the second writing task was significantly higher than those of the control group, and four out of six dimensions were significantly better in the two experimental groups than in the control group.

Lee (2020) conducted a six-week teaching practice, and asked 34 Korean English majors to produce an essay in Korean after watching a 15-minute TED video and then translate the essay into English both manually and automatically. Finally, the students revised their own translations with MT. The results showed that MT helped reduce lexical and grammatical errors in essays, draw students' attention to the writing processes and improve their writing strategies. However, due to lack of a pretest, a post-test, and a control group, it remained unknown whether students' writing competence had really improved and whether it was MT that could make a difference.

C. Students' Stances of MT in Writing Classrooms

Generally, most subjects in the literature feel positive towards integrating MT into L2 writing. For example, most of Lee's (2020) subjects credited MT for its accuracy, provision of authentic expressions and effectiveness for lexical choices. However, some turned up their noses at MT due to its abundant grammatical errors, problematic syntactic structures and awkward literal translations. Moreover, the attitudes of EFL learners with differing writing proficiency levels were mixed. More proficient English learners praised MT's effectiveness at the vocabulary level rather than at the sentence level, while less proficient learners commended MT's usefulness for grammatical and lexical error correction. Bahri and Mahadi (2016) reported on most participants' agreement with using Google Translate for acquiring the writing skills, their slight disagreement with using it for learning grammar and their preference to use it for vocabulary acquisition. Niño (2009) investigated tutors' and learners' perceptions of using MT for Spanish teaching and learning and obtained similar research results.

To date, MT is still criticized for its low accuracy, but language teachers can't ignore students' tremendous demand for its auxiliary functions. Considering this, this study is intended to optimize the research design of the studies in this field and empirically explore the possibility of incorporating MT into EFL writing instruction in mainland China. Specifically, this study tests whether MT can improve learners' writing ability, how EFL learners utilize machine translations to assist their writing processes, and how they respond to such a novel teaching and learning practice.

III. RESEARCH METHODOLOGY

A. Research Questions

(1) During the semester-long teaching experiment, how do EFL learners use MT to assist their writing process?
(2) Can MT improve EFL learners' writing competence?
(3) How do learners perceive integrating MT into EFL writing instruction?

B. Subjects

The subjects came from two intact classes of second-year English majors in a university in Southwest China, and they were divided into an experimental class (6 boys and 36 girls) and a control class (5 boys and 35 girls). Their writing teacher rated them as low-intermediate EFL writers with weak lexical and grammatical foundations. *T*-test shows that no significance existed between both classes' average scores of the course *Basic Writing II* (p>.05). In the fall semester of 2020-2021 academic year, all subjects attended the course *Intermediate English Writing I* lectured by the second author of this article.

C. Experimental Design

This teaching experiment lasted for 16 weeks. The teacher of *Intermediate Writing I* lectured on the same amount of writing knowledge to both classes. During the teaching process, all subjects were required to finish the same topic every three weeks, and complete five writing tasks in total. The third week for each topic was the classroom writing time (90 minutes). The teacher demanded students in both classes to participate in pre-writing discussions, brainstorming and other activities. The differences in writing activities for both classes are displayed in Figure 1.
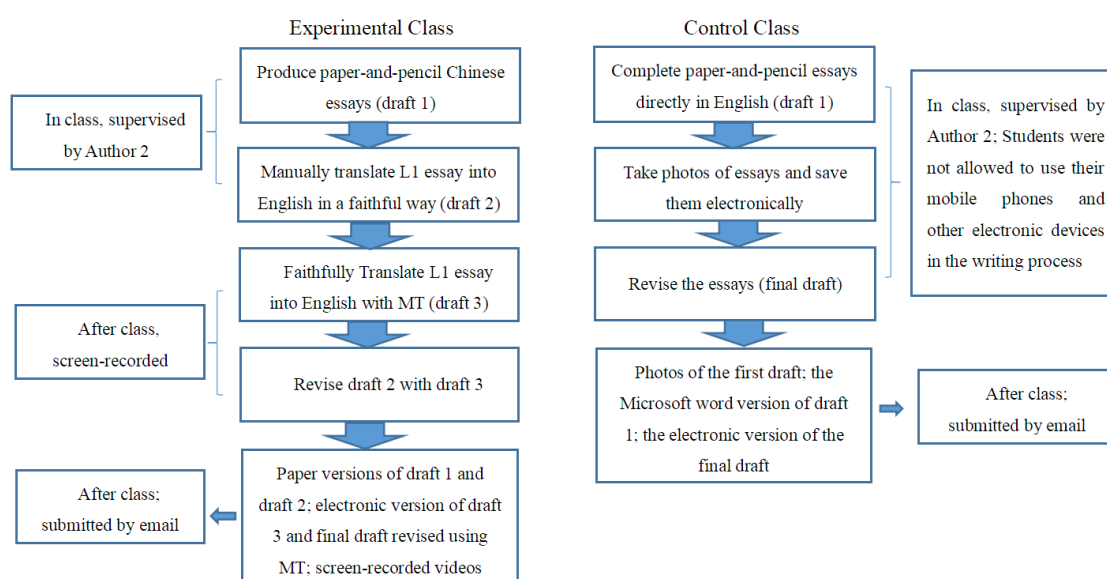

Fig.1 Differences in writing activities for both classes

D. Instruments, Data Collection and Analysis

First, the screen-recording tool *Screencast-O-Matic* was used to keep track of the post-translation editing and revising processes of the experimental class. Based on the classification standard of the revising process proposed by Ferris (1997), we established an analysis framework suitable for this study (see Table 1), including four types (*i.e.* replace, add, delete and transpose) and four levels (*i.e.* writing mechanics, vocabulary & phrase, sentence & grammar and discourse coherence). The second author analyzed the first and final drafts for each task submitted by students in the experimental class and the screen-recorded videos to classify and count the students' post-editing operations in terms of the revision types and levels. The first author did the checking work. Where differences arose, both researchers discussed and negotiated with each other until a consensus was reached.

TABLE 1
EXAMPLES FOR REVISION TYPES AND LEVELS

Revision types	Revision levels	Examples	
		First draft	Final draft
Replace	Mechanics	It is true education will *styfle* creativity...	It is true education will *stifle* creativity...
	Vocabulary	Education make children smarter than...	Education *enables* children *to be* smarter than...
	Sentence	Parents and teachers should guide their children correctly...	It is parents' and teachers' responsibility to guide their children...
	Coherence	Despite it poses threat to one's development...	Although it poses threat to kids' development...
Add	Mechanics	However school education does kills creativity.	However, school education does kill creativity.
	Vocabulary	Parents and teachers should guide their children correctly...	It is parents' and teachers' *responsibility* to guide their children...
	Sentence	No one is supposed ignore the role of education.	No one is supposed *to* ignore the role of education.
	Coherence	It is true education will styfle creativity, we should be cautious about...	It is true education will stifle creativity, *so* we should be cautious about...
Delete	Mechanics	Education is *"key"* to our mental and physical health.	Education is *key* to our mental and physical health.
	Vocabulary	Parents and teachers should guide their children *correctly*...	It is parents' and teachers' responsibility to guide their children...
	Sentence	However school education *does kills* creativity.	However, school education *does kill* creativity.
	Coherence	/	/
Transpose	Mechanics	/	/
	Vocabulary	I *also would like* to hold the idea that...	I *would also like* to hold the idea that...
	Sentence	Only by receiving schooling *kids can* acquire sufficient knowledge necessary in life.	Only by receiving schooling *can kids* acquire sufficient knowledge necessary in life.
	Coherence	/	/

We conducted a pretest and a post-test in week 1 and week 16 respectively. Both tests were of the same topic and required students in both classes to finish an in-class time-limited paper-and-pencil writing task (see Appendix 1). The test results were used to compare the writing scores and the quantitative indices of the written texts prior to and after the experimental teaching. The essays were independently scored by two veteran essay raters based on the rating rubrics (see Appendix 2) of the Test for English Majors Band 4, a standardized test for second-year English majors in mainland China. Any disagreement was resolved through negotiation between both raters and the final score of each essay was the average of the scores assigned by both raters. The agreement between both raters reached 92.68%.

The language accuracy figure of each essay was computed with the formula: 100—(total number of errors ÷ total number of words)×100 (Bai & Ye, 2018). Essay errors were automatically analyzed and counted by Antidote, a kind of software which can identify three types of errors in English and French, namely language use, typography and style. We only calculated errors in language use in the current study. Both authors double-checked the language errors identified by Antidote in the same batch of articles (20 articles in total) one by one to ensure the validity of the errors. The agreement rate of error discrimination between the two authors was 95.67%. Where disagreement occurred, both discussed and solved the differences, and the task involving error discrimination of the remaining 144 essays was completed by the second author.

12 textual quantitative indices in four dimensions were gleaned, including fluency, lexical complexity (Paul, 2005; Lu, 2012), syntactic complexity (Lu & Xu, 2016) and discourse coherence (Jiang, 2016). The principles of selecting indices are economy and representativeness. Detailed information on the indices is listed in Appendix 3.

The research instruments also include an open-ended questionnaire and a quantitative one. The former was mainly designed to see whether the control group would use MT engines while revising their essays after class (see Appendix 4). The latter is a five-point Likert scale with 20 sub-questions concerning the attitudes of the experimental group towards MT's role in EFL writing (see Appendix 5). Both questionnaires were administered in week 16 on the Tencent QQ platform, a widely used social networking medium in mainland China.

SPSS 20.0 was run to compare the scores, accuracy figures and 12 textual quantitative indices of all the essays, and the significance value was set at the p<.05 level. Questionnaire statistics were calculated in Excel worksheets.

IV. RESEARCH RESULTS

A. Results of Students' Writing and Revising Behaviors

Figure 2 shows a Chinese essay and its manual English translation produced by one of the students in the experimental class. As can be seen, this student made quite a lot of self-directed modifications in both versions prior to automatic translation even though such revisions were not required by the teacher.
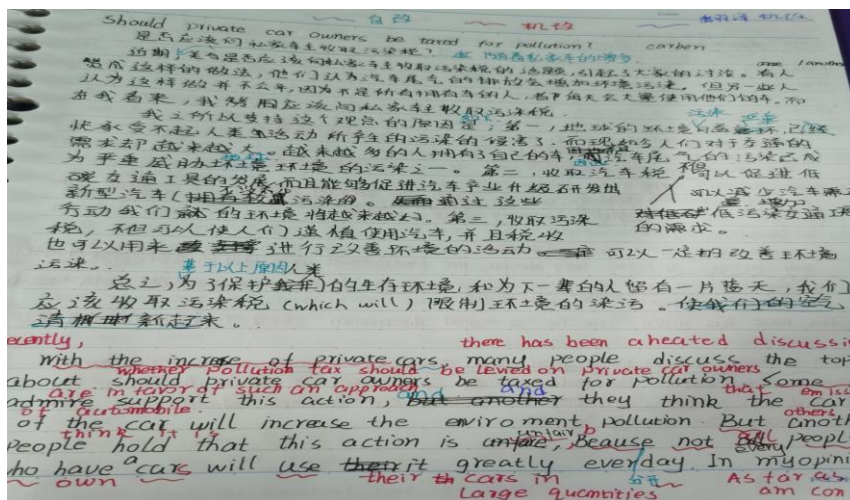
Fig.2 A snapshot of a Chinese essay and its manual translation

Figure 3 displays the interface of Google Translate where the source and target languages are listed in the form of parallel texts.

Fig.3 A screenshot of a Chinese essay and its MT version by Google Translate

Figure 4 presents how students revised their manual translations by using MT versions. Students typically compared the MT version with the manual version and decided whether to make changes. If any change was made, they would highlight it in the form of comments juxtaposed in the margin of a Microsoft word document.

Fig.4 A screenshot of one student's revising process

All together, the experimental class produced 420 essays with 112,140 English words. Undergoing laborious and backbreaking class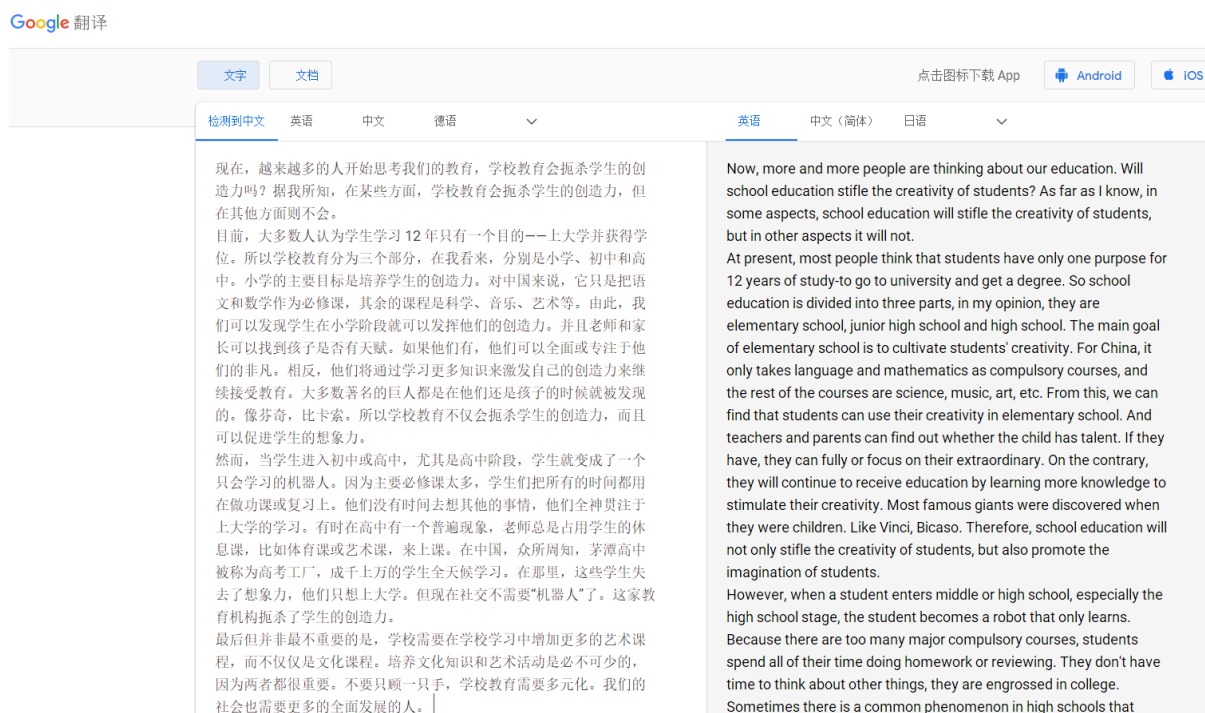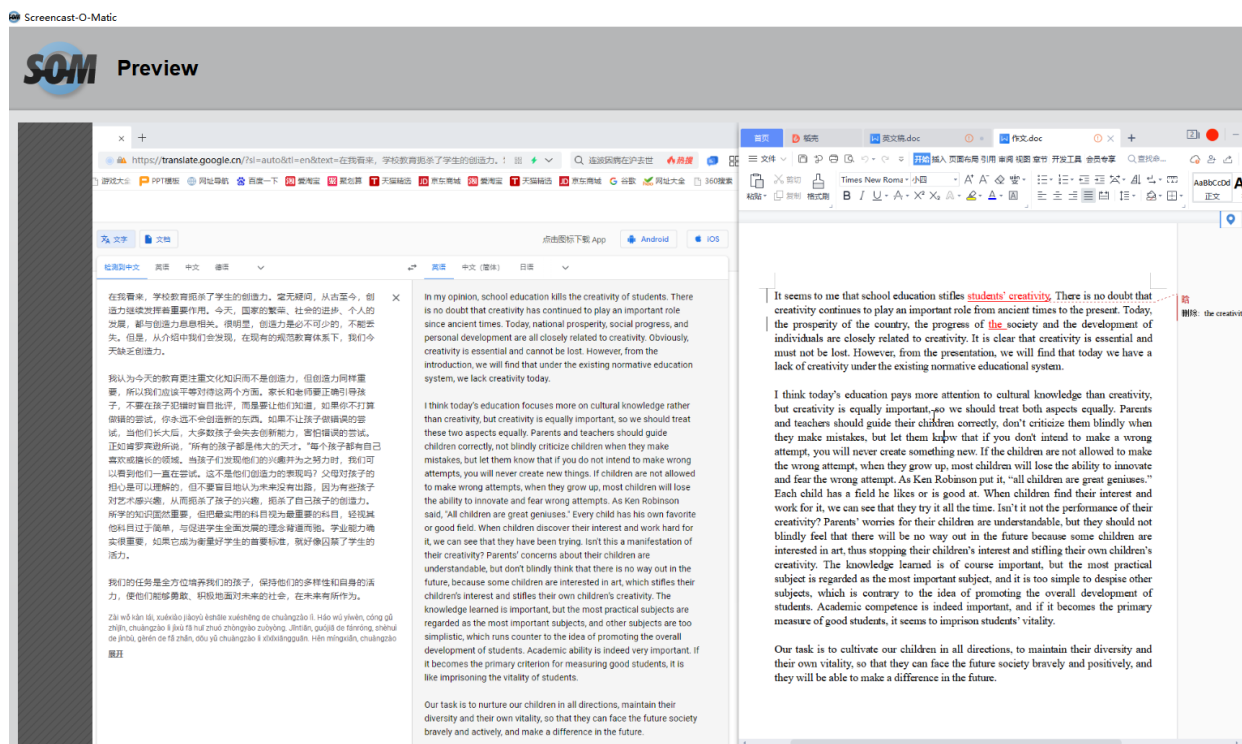ification, confirmation and calculation, both researchers ultimately got a clear picture of how students used MT to revise their essays. The quantitative data relevant to revision types and levels for each task were exhibited in Table 2 which shows that students made more revisions in lexical items than in any other aspects.

TABLE 2
REVISION FIGURES OF THE EXPERIMENTAL CLASS

Revision types	Revision levels	Task 1	Task 2	Task 3	Task 4	Task 5	Total
Replace	Mechanics	423	348	336	214	121	1442
	Vocabulary	838	765	739	678	732	3752
	Sentence	238	217	199	234	179	1067
	Coherence	156	187	164	123	171	801
Add	Mechanics	52	38	54	39	23	206
	Vocabulary	369	421	386	291	265	1732
	Sentence	235	222	312	256	301	1326
	Coherence	253	167	203	194	98	915
Delete	Mechanics	13	21	26	18	19	97
	Vocabulary	206	217	189	217	234	1063
	Sentence	189	176	204	193	213	975
	Coherence	0	0	0	0	0	0
Transpose	Mechanics	0	0	0	0	0	0
	Vocabulary	127	151	111	126	130	645
	Sentence	118	147	108	92	103	568
	Coherence	0	0	0	0	0	0
Total		3217	3077	3031	2675	2589	14589

B. Results of Between-Groups Comparison

Independent-samples *t* test reveals that in the pretest the average score of the control class is significantly higher than that of the experimental class by 4.5375 points (t=-7.483, df =39, p=.000<.05, see Table 3). No significant difference exists in textual quantitative indices (p>.05). Essay accuracy of the control class is significantly higher than that of the experimental class (average difference=-2.725, p=.022<.05).

TABLE 3
BETWEEN-GROUPS COMPARISON OF SCORES AND TEXTUAL QUANTITATIVE INDICES IN THE PRETEST

E-C[1] (I-J)	Average difference(I-J)	Standard deviation	Standard error	95% confidence interval		t	df	Sig. (two-tailed)
				Lower limit	Upper limit			
Score	-4.5375	7.2159	1.1409	-10.8453	-6.2297	-7.483	39	.000
Accuracy	-2.725	7.211	1.140	-5.031	-.419	-2.390	39	.022
LD	.00175	.06300	.00996	-.01840	.02190	.176	39	.861
LV	.01125	.15098	.02387	-.03704	.05954	.471	39	.640
U index	-.43250	3.11577	.49265	-1.42897	.56397	-.878	39	.385
$K2^{+}$	.0025	.5409	.0855	-.1705	.1755	.029	39	.977
W	11.800	42.307	6.689	-1.731	25.331	1.764	39	.086
MTL	-.068925	2.730687	.431759	-.942241	.804391	-.160	39	.874
CT/T	-.0384	.3624	.0573	-.1543	.0775	-.670	39	.507
CRFSO	-.023400	.242428	.038331	-.100932	.054132	-.610	39	.545
CRFAO	.008000	.285200	.045094	-.083211	.099211	.177	39	.860
LSAPP	-.067525	.283047	.044754	-.158048	.022998	-1.509	39	.139
CNCA11	9.96130	41.04553	6.48986	-3.16570	23.08830	1.535	39	.133
WRDPRO	12.7194	91.54078	14.4738	-16.55673	41.99558	.879	39	.385

Table 4 shows that the average score of the experimental class is significantly higher than that of the control class (mean score difference=2.913, p=.027<.05). The accuracy figures and four lexical complexity indices of the experimental class are significantly higher than those of the control class (p<.05).

TABLE 4
BETWEEN-GROUPS COMPARISON OF SCORES AND TEXTUAL QUANTITATIVE INDICES IN THE POST-TEST

E -C (I-J)	Average difference (I-J)	Standard deviation	Standard error	95% confidence interval		t	df	Sig. (two-tailed)
				Lower limit	Upper limit			
Score	2.913	4.701	.743	.416	2.591	1.228	39	.027
Accuracy	7.050	8.539	1.350	4.781	9.319	5.221	39	.000
LD	.02125	.05534	.00875	.00355	.03895	2.429	39	.020
LV	.13550	.13078	.02068	.09368	.17732	6.553	39	.000
U index	4.12825	4.05706	.64148	2.83074	5.42576	6.436	39	.000
$K2^{+}$	3.4650	1.4591	.2307	2.9984	3.9316	15.019	39	.000
W	-.875	72.069	11.395	-23.924	22.174	-.077	39	.939
MTL	-.7364525	4.2545886	.6727095	-2.0971359	.624230	-1.095	39	.280
CT/T	-.0685000	.4757490	.0752225	-.2206519	.0836519	-.911	39	.368
CRFSO	.171425	.335762	.053089	.064043	.278807	3.229	39	.053
CRFAO	.082075	.358601	.056700	-.032611	.196761	1.448	39	.156
LSAPP	-.078800	.258447	.040864	-.161455	.003855	-1.928	39	.061
CNCA11	-1.80320	47.99588	7.58881	-22.15302	8.54662	-.896	39	.375
WRDPRO	1.65312	75.35176	11.9141	19.55446	67.75178	3.664	39	.081

C. Results of Within-Group Comparison

Paired-samples *t* test shows that the post-test scores of the experimental class are significantly higher than the pretest scores (average difference=-12.6786, p=.000), and the total scores of the subjects improved by more than 12 points. Among the quantitative indices, except that all discourse indices have no significant difference (p>.05), the values of the other indices in the post-test are significantly higher than those in the pretest (p<.05).

TABLE 5
WITHIN-GROUP COMPARISON OF SCORES AND TEXTUAL QUANTITATIVE INDICES (THE EXPERIMENTAL CLASS)

Pre-Post （I-J)	Average difference (I-J)	Standard deviation	Standard error	95% confidence interval		t	df	Sig. (two-tailed)
				Lower limit	Upper limit			
Score	-12.6786	6.4220	.9909	-14.6798	-10.6773	-12.795	41	.000
Accuracy	-3.857	7.801	1.204	-1.426	-6.288	-3.204	41	.003
LD	-.02786	.05655	.00873	-.04548	-.01024	-3.193	41	.003
LV	-.15310	.12520	.01932	-.19211	-.11408	-7.925	41	.000
U index	-5.82690	2.72836	.42100	-6.67712	-4.97669	-13.841	41	.000
$K2^{+}$	-8.5024	2.9564	.4562	-9.4237	-7.5811	-18.638	41	.000
W	-65.000	61.743	9.527	-84.241	-45.759	-6.823	41	.000
MTL	-1.493485	3.291461	.507883	-2.519177	-.467794	-2.941	41	.005
CT/T	-.3272952	.4302754	.0663929	-.4613785	-.1932120	-4.930	41	.000
CRFSO	-.208542	.289040	.044600	-.299024	-.118881	-4.685	41	.072
CRFAO	-.085048	.320048	.049385	-.184782	.014686	-1.722	41	.093
LSAPP	.072	.245	.038	-.004	.149	1.917	41	.062
CNCA11	-1.67695	41.05714	6.33525	-.88264	-26.47125	-2.159	41	.087
WRDPRO	-1.96219	79.91670	12.33141	-59.86598	-10.05839	-2.835	41	.067

---

[1] The letter E stands for the experimental class and C for the control class. The two letters have the same meanings in the following tables.

Table 6 shows that the post-test scores of the control class are significantly higher than the pretest scores (average difference=-4.8875, p=.000), but the improvement range is smaller than that of the experimental class. Among the textual quantitative indices, significant differences exist in two lexical complexity indices, fluency and two syntactic complexity indices (p<.05), indicating that subjects in the control class have improved significantly in these aspects.

TABLE 6
WITHIN-GROUP COMPARISON OF SCORES AND TEXTUAL QUANTITATIVE INDICES (THE CONTROL CLASS)

Pre-Post (I-J)	Average difference (I-J)	Standard deviation	Standard error	95% confidence interval		t	df	Sig. (two-tailed)
				Lower limit	Upper limit			
Score	-4.8875	4.1625	.6582	-6.2187	-3.5563	-7.426	39	.000
Accuracy	-.725	10.639	1.682	-4.127	2.677	-.431	39	.669
LD	-.00700	.05170	.00817	-.02353	.00953	-.856	39	.397
LV	-.03050	.15643	.02473	-.08053	.01953	-1.233	39	.225
U index	-1.30000	3.51222	.55533	-2.42326	-.17674	-2.341	39	.024
$K2^+$	-5.1925	1.8199	.2877	-5.7745	-4.6105	-18.045	39	.000
W	-76.425	63.093	9.976	-96.603	-56.247	-7.661	39	.000
MTL	-2.055857	3.517509	.556167	-3.180811	-.930903	-3.696	39	.001
CT/T	-.3307275	.5124968	.0810329	-.4946319	-.1668231	-4.081	39	.000
CRFSO	.002400	.274648	.043426	-.085437	.090237	.055	39	.956
CRFAO	.001825	.270881	.042830	-.084807	.088457	.043	39	.966
LSAPP	.040800	.296440	.046871	-.054006	.135606	.870	39	.389
CNCA11	-2.84767	44.36505	7.01473	-17.03630	11.34095	-.406	39	.687
WRDPRO	1.321625	79.925847	12.637386	-24.239901	26.883151	.105	39	.917

D. Results from the Questionnaires

The results of the open-ended questionnaire show that 12.5% (n=5) of the students in the control class used MT engines in essay revision after class, and among them 80% (n=4) consulted *Youdao Dictionary* (an online dictionary which has both a web version and a mobile version) for translating some uncertain expressions or confirming lexical forms and usages. When looking up a word, 60% (n=3) tended to notice the synonymous words with lower frequency and use them to replace high-frequency words to increase the lexical complexity of their essays. They believed that by referring to the sample sentences they could accumulate advanced vocabulary and use words correctly. In short, only a small proportion of students in the control class utilized MT in the revising and editing process, which ensures the validity and comparability of the quantitative data.

Results of the quantitative questionnaire are set out in Table 7. Questions 10, 13, 15 and 19 surveyed the experimental subjects' overall understanding of MT and its integration into the writing process. The results show that most of the 42 students held a positive attitude towards the accuracy of MT (69.01%, n=29), and its helpfulness in improving the writing proficiency (71.43%, n=30). 88.10% (n=37) complimented the "writing in Chinese first—translating manually—translating automatically—comparing both translations—revising" process, and 90.48% (n=38) claimed that they would continue to use MT to assist the writing process.

TABLE 7
DESCRIPTIVE STATISTICS FOR THE ATTITUDE QUESTIONNAIRE[2]

Dimension	Item	N	Min.	Max.	Mean	Std. Deviation	Overall attitude
Overall understanding	10	42	1	5	4	1.082	Disagree
	13	42	2	5	4.07	1.113	
	15	42	1	5	1.43	0.914	
	19	42	1	4	1.33	0.721	Strongly agree
Attitude towards MT's learning -promoting effects	1	42	1	4	1.64	1.122	
	2	42	1	5	2.05	1.413	Agree
	3	42	1	4	1.64	1.144	
	4	42	1	4	1.48	0.862	Strongly agree
	5	42	1	5	1.52	1.042	
	6	42	1	5	1.86	1.221	
	7	42	1	5	1.9	1.284	Agree
	9	42	1	5	2	1.342	
	12	42	1	5	2.14	1.28	
	14	42	1	3	1.31	0.643	
	17	42	1	4	1.55	0.993	Strongly agree
Use of MT in the writing process	11	42	1	3	1.38	0.731	
	18	42	1	5	2.6	1.251	
	20	42	1	5	2.95	1.447	Slightly disagree
Others	8	42	1	5	3.17	1.342	
	16	42	1	5	1.6	1.061	Strongly agree

Questions 1 to 7 and 9, 12, 14, and 17 are pertinent to MT's learning-promoting effects. Most of the subjects said that MT helped to organize their ideas (73.81%, n=31), raise their awareness of text organization (61.90%, n=26), make their essays more fluent (76.19%, n=32), reduce mechanical errors (80.95%, n=34), lexical errors (85.71%, n=36) and grammatical errors (69.05%, n=29), improve lexical complexity (69.05%, n=29) and syntactic complexity (59.52%, n=25), accumulate authentic expressions (52.38%, n=22), and reduce the anxiety and nervousness they experienced in the course of writing (90.48%, n=38), which in turn enhanced their self-confidence and self-efficacy in English writing (80.95%, n=34).

Questions 11, 18, and 20 address how subjects in the experimental class utilized MT. More than 80% of the subjects used MT to correct lexical errors (85.71%, n=36) and 45.24% (n=19) corrected grammatical errors, but less than 40% employed MT to optimize the text organization (38.10%, n=16), although more than 60% believed that the incorporation of MT would help improve the awareness of text organization.

In addition, it is worth mentioning that although 83.33% (n=35) indicated that they could distinguish the difference between MT and manual translation, only 28.56% (n=12) claimed that they did not encounter any technical problems while applying MT engines, which requires writing teachers to provide adequate training for students to ensure smooth progress of the teaching process.

V. DISCUSSION

A. EFL learners' Use of MT in the Writing Process

Integrating MT in EFL writing is a novel instructional attempt in mainland China. In the current study, students were required to utilize MT in the writing process and the results revealed that students paid greater heed to the lexical level of all four revision types whether it involves replacing, adding, deleting or transposing a certain lexical item. This is explicable probably because constrained by their English proficiency low-intermediate EFL learners are more likely to notice lexical differences between two parallel texts than complicated syntactic patterns, cohesive devices or other demanding aspects. This finding is not in accordance with that of Lee's (2020) study where the subjects were more open to making grammatical changes. A robust explanation of this discrepancy might be again of relevance to the difference in the English proficiency of both student populations. The English level of Lee's (2020) subjects was between intermediate and high-intermediate, while that of the subjects in the current study was between low and intermediate. EFL learners at the intermediate and lower levels are often confusedly challenged by the lexical and grammatical errors dotted in their essays which often impede communication and lower the quality of their writing (Lee, 2014). However, compared with grammatical errors or structures, lexical items are more noticeable when two parallel texts are put side by side, probably because identification of problematic grammars may involve greater cognitive load on the part of low-proficiency learners.

---

[2] Score 1 denotes "totally agree" and score 5 "totally disagree".

Within-group comparisons might also triangulate the revision figures, which show that the experimental class significantly improved their final version through using MT. This improvement can be proved by the fact that students made more lexical and grammatical revisions in the writing process, presumably highlighting the accuracy at both levels. Previous research concluded that improved grammar is characteristic of enhanced text quality (Min, 2006) and that grammar correction promotes the communicative effectiveness of writing (Rahimi, 2009). Although lexico-grammatical accuracy is not a sole benchmark for good English writing, it can surely contribute to the quality of writing. That's why students paid more attention to these two aspects than others.

B. Impacts of MT on EFL Learners' Writing Competence

The comparisons of essay scores between groups show that the pretest scores of the control class are significantly higher than those of the experimental class, while the results are opposite in the post-test. Comparisons within the group show that the post-test scores of the experimental class are significantly higher than those of the pretest scores, with a mean difference of 12 points. The post-test scores of the control class are also significantly higher than those of the pretest, but the mean difference is less than 5 points. It can be seen that, after one semester of study, the overall writing performance of the experimental class got much better than that of the control class. The courses completed by the two classes in one semester were almost identical, and the writing course was also undertaken by the same teacher. Therefore, the difference in post-test scores between the two classes is likely to be closely related to MT's integration into the writing process. Presumably we can draw a conclusion with a certain confidence that MT engines as CALL tools can improve the overall writing ability of EFL learners. The results of this study are consistent with those of Garcia and Pena (2011), O'Neill (2012), O'Brien (2018) and Lee (2020), but these studies are not entirely aimed at EFL learners, with the first two studies taking Spanish and French learners as their subjects respectively.

Now that MT is instrumental in enhancing learners' overall writing proficiency, in what aspects can its auxiliary function be highlighted? In order to clarify this question, this study made a quantitative analysis of the textual features of all compositions produced by students in both classes. Statistical analyses show that, in terms of accuracy, the experimental class performed significantly worse than the control class in the pretest, but did significantly better in the post-test. Within-group comparisons found that the experimental class did significantly better in the post-test than in the pretest, while for the control class there existed no significant difference. It shows that the accuracy of essays in the experimental class underwent substantial amelioration. As for lexical complexity, there is no significant difference between the two classes in the pretest, and the experimental class did significantly better than the control class in the post-test. The within-group comparisons show that the lexical complexity indices of the experimental class increased significantly, while only the U index and K2+ indices of the control class witnessed a significant increase. The results of accuracy and lexical complexity are the same as those of most studies in this field, that is, MT helps to reduce linguistic errors and enrich the lexical knowledge of language learners (Amaral & Meurers, 2011).

At the fluency level, no significant difference arose between the experimental class and the control class in both pretest and post-test. The essays produced by both classes in the post-test were significantly longer than those produced in the pretest, with an average difference of 65 words for the experimental class and 76.425 words for the control class respectively. The fluency of both classes has improved, but the improvement of the control class is more obvious than that of the experimental class. The improvement of fluency might be a natural growth in the teaching process, which might not have direct relation with the auxiliary function of MT. The results of fluency index in this study are inconsistent with those of Garcia and Pena (2011), which indicated that MT could especially help beginners to communicate more. In the present study, the growth trend of syntactic complexity in both classes is similar to the fluency index, which is consistent with the results of Garcia and Pena (2011), which pointed out that MT could not make the sentence structures more complicated due to the low language level and the low rate of syntactic modifications by learners. Therefore, it is likely that the differences in the results of the two studies might be ascribed to the asymmetry in subjects' language proficiency. The subjects in our study are low-intermediate English learners, while the subjects in Garcia and Pena (2011) were Spanish beginners. However, whether learners' language proficiency will affect the learning-promoting effect of MT needs further confirmation in future research.

At the coherence level, there is no significant difference within and between groups in the pretest and post-test. This can be illustrated by the low rate of revisions at this level in the writing process. The results of this study are consistent with those in the existing studies, some of which pointed out that MT was not helpful to learners in improving the coherence of their essays (Groves & Mundt, 2015).

To sum up, the improvement in scores of the experimental class is mainly due to the improvement of essay accuracy and lexical complexity. Integrating MT into EFL writing process is especially helpful to reduce language errors in students' essays and enables students to accumulate and use more complex vocabulary.

C. EFL Learners' Perceptions of MT in English Writing Instruction

The questionnaire data prove that the students in the experimental class have a generally positive attitude towards the integration of MT into EFL writing instruction. They believed that the process of MT-assisted writing could improve their overall writing ability, and especially help consolidate their lexical and grammatical foundations. This results obtained from the questionnaires largely confirm those displayed in Table 3 to Table 6, and are equally consistent with the results of Lee (2020) which revealed that most subjects believed the effectiveness of MT in improving their overall

writing competence and helping them properly use words and authentic expressions in a specific context, and that a majority of the subjects would continue to use MT to locate words suitable in diverse contexts and raise their meta-linguistic awareness of vocabulary and grammar by comparing the differences between manual and automatic translations. In the emotional dimension, the subjects in the current study believed that MT-assisted writing could reduce their anxiety in the writing process and enhance their self-confidence and self-efficacy in English writing, which is in agreement with the research results of Bahri and Mahadi (2016) and Niño (2009). A rational explanation might be that integrating MT into the writing process presumably frees students from bearing huge cognitive load in the writing process, especially for students whose language proficiency remains at an intermediate or lower level, and thus the advantages of MT in this regard appear more prominent.

VI. PEDAGOGICAL IMPLICATIONS & CONCLUSION

The research results have implications for language teaching. This study finds that MT can be used as an effective teaching aid, so EFL teachers can integrate it into writing instruction and explore its potentials as a CALL tool. As Correa (2014) pointed out, MT technology is constantly improving, and the quality of translation is getting increasingly higher. Against the background of growing demand for MT, EFL writing teachers need to accept its existence reasonably, rather than completely ignore it. However, EFL teachers should pay attention to the fact that no tool serves as a panacea for language teaching. The original intention of designing translation engines is not to promote language learning. In order to benefit learners from MT, teachers must put it in an appropriate position in language learning, instead of relying entirely on it, and give students sufficient guidance on how best to realize its fullest potentials, and inform learners of its strengths and weaknesses (Bahri & Mahadi, 2016).

The current study is significant in that it empirically examined three dimensions, namely the process, product and perceptions related to the utilization of MT in EFL writing instruction. However, the number of subjects is small, which may undermine the generalization of the results, so it is better for future researchers to carry out empirical studies involving different student populations. What deserves a special mention is that the influence of other factors on writing results is not considered. For example, the use of different translation engines and the difference in individual EFL learner's language level are not taken into account. Furthermore, the experimental environment may also affect the research results. Students in the experimental class first finished the paper-and-pencil writing tasks in class, and then completed manual and machine translations, compared both versions and revised the written products on the computer, while students in the control class only finished the paper-and-pencil writing tasks. If all the subjects finished the writing tasks in the laboratory environment, it would be unclear whether the research results will be consistent with the results obtained in the present study. Finally, in view of the popularity of AWE (automated writing evaluation) systems in EFL writing classrooms, comparisons can be made between the effects of AWE systems and MT engines as CALL tools on EFL writing instruction, or their synergistic effects on learners' writing ability can be elucidated.

APPENDIX

1. Pretest and post-test writing prompt

Should young criminals be treated in the same way as adults by authorities? This issue has been intensely discussed for years. Give your views on the issue in about 200 words. Marks will be awarded for content relevance, content sufficiency, organization and language quality.

2. Rating rubrics for TEM-4 writing

Score band	Descriptors
14 points	Fully achieves relevance to the topic; clearly presents ideas; skillfully manages cohesion and coherence; uses a wide range of sophisticated lexical and syntactic structures with basically no language errors.
11 points	Achieves great relevance to the topic; clearly presents ideas; uses a range of cohesive devices although there may be some over/under-use; uses complex lexical and syntactic structures with occasional minor language errors.
8 points	Achieves relevance to the topic; generally presents ideas in a clear and coherent way; attempts to use uncommon lexical items and complex syntactic structures although there are many language errors, including a few serious ones which may distort the message.
5 points	Basically achieves relevance to the topic; fails to clearly present ideas in a clear enough way; fails to make the whole essay a coherent one; uses a wide range of high-frequency lexical items and simple and fragmented syntactic structures; commits many serious language errors which distort the message.
2 points	Does not achieve relevance to the topic; presents ideas illogically and incoherently; uses high-frequency lexical items and cannot use sentence forms; commits a wide range of serious mistakes which distort the message.
0 point	Fails to communicate any message relevant to the topic, or writes only a few words, or writes a completely memorized response.

3. Information on textual indices

Dimension	Indices	Code	Computing tool
Lexical complexity	Lexical Diversity	LD	L2 Lexical Complexity Analyzer
	Lexical Variation	LV	
	Uber Index	U	
	Words beyond the frequency of 1-2000	K2$^+$	Vocabprofile
Fluency	Essay Length	W	L2 Syntactic Complexity Analyzer
Syntactic complexity	Mean T-unit Length	MTL	
	Ration of Complex T-unit	CT/T	
Discourse coherence	Stem Overlap in Adjacent Sentence	CRFSO	Coh-Metrix 3.0
	Argument Overlap in Adjacent Sentences	CRFAO	
	LSA Overlap in Adjacent Paragraphs	LSAPP	
	All Connective Incidence	CNCALL	
	Pronoun Incidence	WRDPRO	

4. Open-ended questionnaire (translated from Chinese)
    1. Do you use any tools to revise your essays after class?
    2. Do you think these tools can help you improve your writing competence? If so, in what ways?
5. Attitude questionnaire (translated from Chinese)

This is a questionnaire concerning your views on incorporating MT into English writing. Tick *only one choice* in the bank for each question. There are no right or wrong answers. We appreciate your honest answers. Thank you!	Strongly disagree	agree	uncertain	disagree	Strongly disagree
1. Writing in Chinese prior to manual translation helps organize ideas.					
2. Revising manual translation using MT raises textual organization awareness.					
3. Referring to MT while revising essays helps reduce mechanical errors.					
4. Writing with MT reduces anxiety and nervousness in the writing process.					
5. Revising essays with MT reduces lexical errors in writing.					
6. Writing with MT enhances my confidence in the writing process.					
7. Revising essays with MT reduces grammatical errors in writing.					
8. I meet no technical problems in using MT engines.					
9. Using MT as an auxiliary tool enables me to write more fluently.					
10. MT is inaccurate and cannot help with my English writing.					
11. I often refer to MT to revise lexical errors in my own translation.					
12. By referring to MT I can use complicated words properly.					
13. Referring to MT does little help to improving my writing ability.					
14. Referring to MT helps me write complicated sentence structures properly.					
15. The process of "writing Chinese—manually translating—MT—comparing manual and MT versions—revising" helps improve my writing ability.					
16. I can tell the difference between manual translations and MT versions.					
17. I can learn many authentic expressions from MT versions.					
18. I often refer to MT to revise grammatical errors in writing.					
19. I will use MT engines to help with writing in the future.					
20. I often consult MT versions to correct coherence problems in manual translations.					

REFERENCES

[1]   Alhaisoni, E. & Alhasysony, M. (2017). An investigation of Saudi EFL University students' attitudes towards the use of Google translate. *International Journal of English Language Education*, *5*(1), 72–82.
[2]   Amaral, L. A. & Meurers, D. (2011). On using intelligent computer-assisted language learning in real-life foreign language teaching and learning. *ReCALL*, *23*(1), 4–24.
[3]   Bahri, H. & Mahadi, T. (2016). Google Translate as a supplementary tool for learning Malay: A case study at Universiti Sains Malaysia. *Advances in Language and Literary Studies*, *7*(3), 161–167.
[4]   Bai, L. F., Ye, S. F. (2018). The dynamic development of English-L2 writing competence. *Modern Foreign Languages*, *41*(3), 354–366.
[5]   Baraniello, V., Degano, C., Laura, L., et al. (2016). A wiki-based approach to computer-assisted translation for collaborative language learning (pp.369–379). In Li, Y., Chang, M., Kravcik, M., *et al.* (eds). *State-of-the-Art and Future Directions of Smart Learning*. Singapore: Springer.
[6]   Bernardini, S. (2016). Discovery learning in the language-for-translation classroom: corpora as learning aids. *Cadernos de Tradução*, *36*(spe), 14–35.
[7]   Carroll, S. & Swain, M. (1993). Explicit and implicit negative feedback: An empirical study of linguistic generalisations. *Studies in Second Language Acquisition*, *15*(3), 357–386.
[8]   Chéragui, M. A. (2012). Theoretical overview of machine translation. *Proceedings ICWIT,* 160–169. Retrieved February 25, 2021, from http://ceur-ws.org/Vol-867/Paper17.pdf

[9]   Correa, M. (2014). Leaving the "peer" out of peer-editing: Online translators as a pedagogical tool in the Spanish as a second language classroom. *Latin American Journal of Content & Language Integrated Learning*, *7*(1), 1–20.

[10]   Ferris, D.R. (1997). The influence of teacher commentary on student revision. *TESOL Quarterly*, *31*, 315–339.

[11]   Garcia, I. & Pena, M. (2011). Machine translation-assisted language learning: writing for beginners. *Computer Assisted Language Learning*, *24*(5), 471–487.

[12]   Godwin-Jones, R. (2015). Contributing, creating, curating: Digital literacies for language learners. *Language Learning & Technology*, *19*(3), 8–20.

[13]   Groves, M. & Mundt, K. (2015). Friend or foe? Google translate in language for academic purposes. *English for Specific Purposes*, *37*(1), 112–121.

[14]   Jiang, J. L. (2016). The application of Coh-Metrix in foreign language teaching and research. *Foreign Languages in China*, *13*(5), 58–65.

[15]   Kliffer, M. (2005). An experiment in MT post-editing by a class of intermediate/advanced French majors (pp.160–165). In *Proceedings of the 10th EAMT Conference: Practical applications of machine translation*. Retrieved May 15, 2008, from https://aclanthology.org/2005.eamt-1.22/

[16]   Lee, S. M. (2014). Grammatical Error Patterns in EFL Students' Writing across Different Language Proficiency Levels. [*Foreign language education*], *21*, 17–36.

[17]   Lee, S. M. (2020). The impact of using machine translation on EFL students' writing. *Computer Assisted Language Learning*, *33*(3), 157–175.

[18]   Lu, X. F. (2012). The relationship of lexical richness to the quality of ESL learners' oral narratives. *Modern Language Journal,* *96*(2), 190–208..

[19]   Lu, X. F. & Xu, Q. (2016). L2 syntactic complexity analyzer and its applications in L2 writing research. *Foreign Language Teaching and Research*, *48*(3), 409–420+479–480.

[20]   Min, H. T. (2006). The effects of trained peer review on EFL students' revision types and writing quality. *Journal of Second Language Writing*, *15*(2), 118–141.

[21]   Niño, A. (2009). Machine translation in foreign language learning: Language learners' and tutors' perceptions of its advantages and disadvantages. *ReCALL*, *21*(2), 241–258.

[22]   O'Brien, S., Michel, S. & Marie-Josee, G. (2018). Machine translation and self-post-editing for academic writing support: Quality explorations (pp.237–262). In Moorkens, J., Castilho, S., Gaspari, F., *et al*. (eds) *Translation quality assessment, machine translation: technologies and applications*. Cham: Springer.

[23]   O'Neill, E. (2012). *The Effect of Online Translators on L2 Writing in French*. Urbana-Champaign: University of Illinois.

[24]   Paul, M. (2005). Lexical frequency profiles: A Monte Carlo analysis. *Applied linguistics*, *26*(1), 32–47.

[25]   Rahimi, M. (2009). The role of teacher's corrective feedback in improving Iranian EFL learners' writing accuracy over time: is learner's mother tongue relevant? *Reading & Writing*, *22*(2), 219–243.

[26]   Wong, T. & Lee, J. (2016). Corpus-based learning of Cantonese for Mandarin speakers. *ReCALL*, *28*(2), 187–206.

Jian Wang was born in Guang'an city, Sichuan Province, China. He received his master's degree in English linguistics from the School of Foreign Languages, Hainan University, China in 2015.

He is currently a lecturer in the School of Humanities, Geely University of China, Chengdu, China. His research interests include EFL teaching, Systemic-Functional Linguistics, and discourse analysis.

Mr. Wang has published several research papers in international and Chinese journals, such as the following publications: Wang, J. & Bai, L.F. (2021). Unveiling the Scoring Validity of Two Chinese Automated Writing Evaluation Systems: A Quantitative Study. *International Journal of English Linguistics*, *11*(2), 68-84; Wang, J. & Zhang, T.Y. (2021). Empirical research on Automated Writing Evaluation systems in China: review and prospect. *Journal of Henan Polytechnic University (Social Sciences)*, *22*(5), 59-66+81; Wang, J. & Wang, X.F. (2020). An analysis of the predictive power of quantitative lexical features on machine-graded essay scores. *Language Education*, *8*(3): 26-32+39; Wang, J. & Zhang, T.Y. (2020). A study of the relationship between the textual quantitative indices of L2 writing and machine scores. *Foreign Language Testing and Teaching*, *3*, 12-20; Bai, L.F. & Wang, J. (2019). A critical review of the effectiveness of automated feedback over the past 20 years. *Foreign Languages Research*, *1*, 65-71+88; Bai, L.F. & Wang, J. (2018). Difference between human and machine scoring and its underlying causes. *Foreign Language Testing and Teaching*, *3*, 44-54.

Xinli Ke was born in Nanchong city, Sichuan Province, China. She received her master's degree in foreign linguistics and applied linguistics from the School of Foreign Languages, Southwest Jiaotong University, China in 2013.

She is currently a lecturer in Department of Foreign Languages, Southwest Jiaotong University Hope College. She is interested in EFL teaching research.