

AI-Assisted Writing Evaluation in an EFL setting: Lessons from China

基于人工智能的写作测评在教学中的应用：来自中国的启示

TANG, Jinlan 唐锦兰

Beijing Foreign Studies University 北京外国语大学

LANGUAGE LEARNING IN THE DIGITAL AGE / 数字时代的语言学习
法国驻华大使馆文化教育合作处、德国歌德学院联合主办
2020年9月4日-6日

北京外国语大学 Beijing Foreign Studies University (BFSU, www.bfsu.edu.cn)

400
Ambassadors

101
languages

Cradle of Diplomats 外交官的摇篮。

Wherever a five-starred red flag is flying, there are Beiwai alumni.
凡是有五星红旗飘扬的地方，就有北外人的身影。

One of the top universities in foreign language studies in China
首批“211工程”高校、“985”优势学科创新平台高校、首批
“双一流”建设高校。



Institute of Online Education, BFSU (BeiwaiOnline) 北京外国语大学网络教育学院 (www.beiwaionline.com)

- Established in December 2000
- **Degree programs**: BA in English, management, computer science and international relations (**10 BAs**) MA in English Language Teaching and in Chinese Language Teaching (**2 MAs**)
- **Staff**: 140+ headquarter staff, 400 local tutors, 190 learning centers
- **Research**: Secretariat for China Computer-Assisted language Learning Association (ChinaCALL, www.chinacall.org.cn)



Log-in page of BeiwaiOnline Learning Platform



Group photo of GloCALL and ChinaCALL 2018

Artificial Intelligence and Human Languages Lab, BFSU 北京外国语大学人工智能与人类语言重点实验室 (ai.bfsu.edu.cn)

- Established in December 2019
- **Mission**: exploring the potential of AI and decoding the mystery of human languages
- **Areas of research**: multilingual brain science; language teaching and learning assisted by technology; multilingual corpus; multilingual natural language processing; multilingualism, cognition and ageing



Lab founding ceremony (Dec, 26th, 2019)



BFSU AI and Human Languages Lab

Outline:

1. Background
2. AI-assisted writing evaluation in action
3. Lessons

1. Background

Challenges of writing instruction:

Formative assessment and feedback of multiple drafts is the key.

- Time-consuming
- Writing expertise
- Feedback ignored and not used

The solution?

Innovations that include strengthening the practice of formative assessment produce significant and often substantial learning gains.

—*Black & William, 1998, p. 140*

AI-assisted writing evaluation

- AI-assisted writing evaluation or Automated writing evaluation (AWE) is a **process of scoring and evaluating written work automatically.**
- **Its origin** can be traced back to the 1960s in the United States with the development of Page Essay Grade (PEG), an e-program that applied multiple regression analysis of measurable features of text, e.g. the average sentence length, to develop a scoring model based on a corpus of essays previously graded manually.
- With the development of **artificial intelligence and natural language processing technologies**, and **the global emphasis on writing instruction**, the last three decades have witnessed rapid development and increasing use of **AWE software in grading standardized tests** such as TOEFL and GMAT and as **a formative assessment tool in L1 and L2 writing.**

Some AI-assisted writing evaluation tools

- WriteToLearn (Pearson)
- iWrite 爱写作 (外研社)

Pearson Professional Large Scale Admissions | Order Status Quick Order Search

Products Featured Topics Digital Solutions Training Order

Home > Professional Assessments > Products > WriteToLearn

WriteToLearn

WriteToLearn is a web-based tool for building writing skills and comprehension in grades 4-12

Overview Benefits Resources Standards Contact Us

iWrite 爱写作 英语写作教学与评阅系统

产品介绍 批改体验 申请试用 关于我们 注册 登录

语言 ★★★★★
内容 ★★★★★
篇章结构 ★★★★★
技术规范 ★★★★★

74.9 总分

评语:
文章的单词拼写错误较少, 句子非常流畅, 能够熟练使用从句, 句式丰富。语法错误较少, 语法结构掌握比较好。高级词汇掌握度不高。内容充实, 篇幅适中, 内容有衔接。

语言 内容

全部 14

词法错误 12

主谓一致错误 1
动词时态错误 1
动词缺失 2
限定词缺失 1
代词缺失 1
名词错用 1
词法错误 1

The advantages of being careful with money in daily life

Nowadays, there **has** 注 a hot debate on whether people should **careful** 注 with their money in their daily life since they try their best to save money. Some people think it is not good to do so because this may not **makes** 注 people smart consumers. And in my eyes, the advantages of being careful with money in daily life **outweighs** 注 the disadvantages.

First, being careful with money is a virtue that can always benefit people. Being thrifty is of great help, especially when the economic **statue** 注 around the world is not very good.

Being money-conscious is one of the good habits that always **helps** 注 people when faced with all kinds of difficult financial **situation** 注. People **have** 注 this awareness will think

2. AI-assisted writing evaluation in action

Research Questions

1. How does the use of AI-Assisted writing evaluation as a formative assessment tool impact college students' **writing performance**?
2. How does it impact the writing process and **learners**?
3. How does it impact the writing teaching process and **teachers**?

Theoretical underpinnings

- **Social informatics theoretical approach** toward the use of AWE, assuming technologies, people, and organizations as a “heterogeneous socio-technical network” (Kling, 1999).
- **Participatory design (PD)** commonly used in human-computer educational research engaging the users of computer systems in designing and revising the computer systems (Steen, 2013)
- **Exploratory practice (EP)** a practitioner-based research combining research and classroom teaching in the natural setting with the aim to resolve teacher and students’ “puzzles” or “problems” in the classroom (Allwright, 2003),

Research Methodology

Participants:

- 5 universities
- 460 students
- 7 teachers

Research Methods:

- Questionnaires
- Interviews
- Journals

Research Interventions:

- **Teaching experiment** integrating **an AI-assisted writing evaluation tool** *Writing Roadmap (WRM)* for 2 semesters
- Teacher support

The AI-assisted writing evaluation tool

-The Writing Roadmap (WRM)

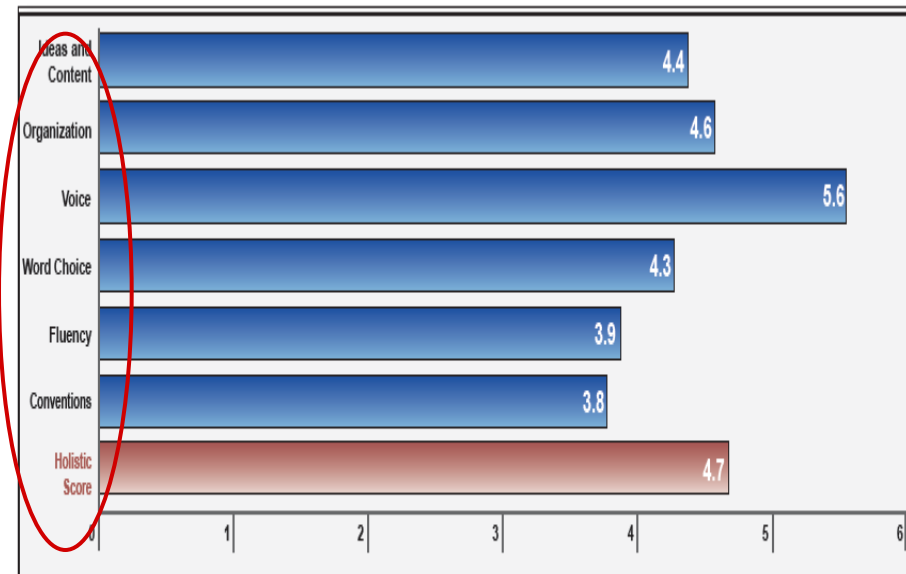
It can provide writing practice and automatic feedback for six writing dimensions and a holistic score for student essays, available for Grades 3-12, college and adult students.

Immediate online feedback through

- highlighting problematic parts
- narrative/comments
- discrete (for each trait) and holistic scores
- remarking and rescoring on revised versions

The screenshot shows the Writing Roadmap (WRM) web interface. The page title is "Enter Essay" and the prompt is "A new student will be coming to your school. Write a letter to him or her, explaining in detail what to expect on the first day of school." The student's title is "The New kid In School". The introduction section contains the text: "Hey, I hear you're new. Don't worry, you're in good hands. My name is Chelsea, and I'm a freshman here at Clark High. I've got some ideas to help you survive this crazy high school. First of all, if you are planning to play a sport, grades are important. You must be passing to be eligible, so reserve time to study and do your homework. Lunch is an hour so you will have time to do extra work then. Second, avoid the last hall in A building. That's where Mrs. Garcia stands and catches you doing anything wrong." The body section contains the text: "I suggests that you hang around with the right people, make good choices and you should not have any problems. Mr. Lamprey is an awesome geography honors teacher. Hopefully, I will get him. Maybe you'll be in some of my classes! If you are, you and I can study together. His tests seem to be pretty difficult, but it's all about knowing the material and how to study." The conclusion section contains the text: "There are many restaurants at which to eat, and you'll make friends easily. Cell phones is allowed of There is a student store in building A. It has a lot of good stuff for sale like candy bars and stacks." A suggestion box is open, showing a redaction of the word "suggests" and a suggestion to use "suggest". The suggestion text reads: "The word (suggests) is a verb in the third person singular. Check its agreement in person with its subject. Rule 2: the verb agrees with the subject(s) in number and in person. Counter examples: he love his cat / they wants more / you was right. Correct examples: he loves his cat / they want more / you were right Suggestion: Change (suggests) to suggest." The interface also includes buttons for "Hint", "Tutor", "Thesaurus", and "Tree".

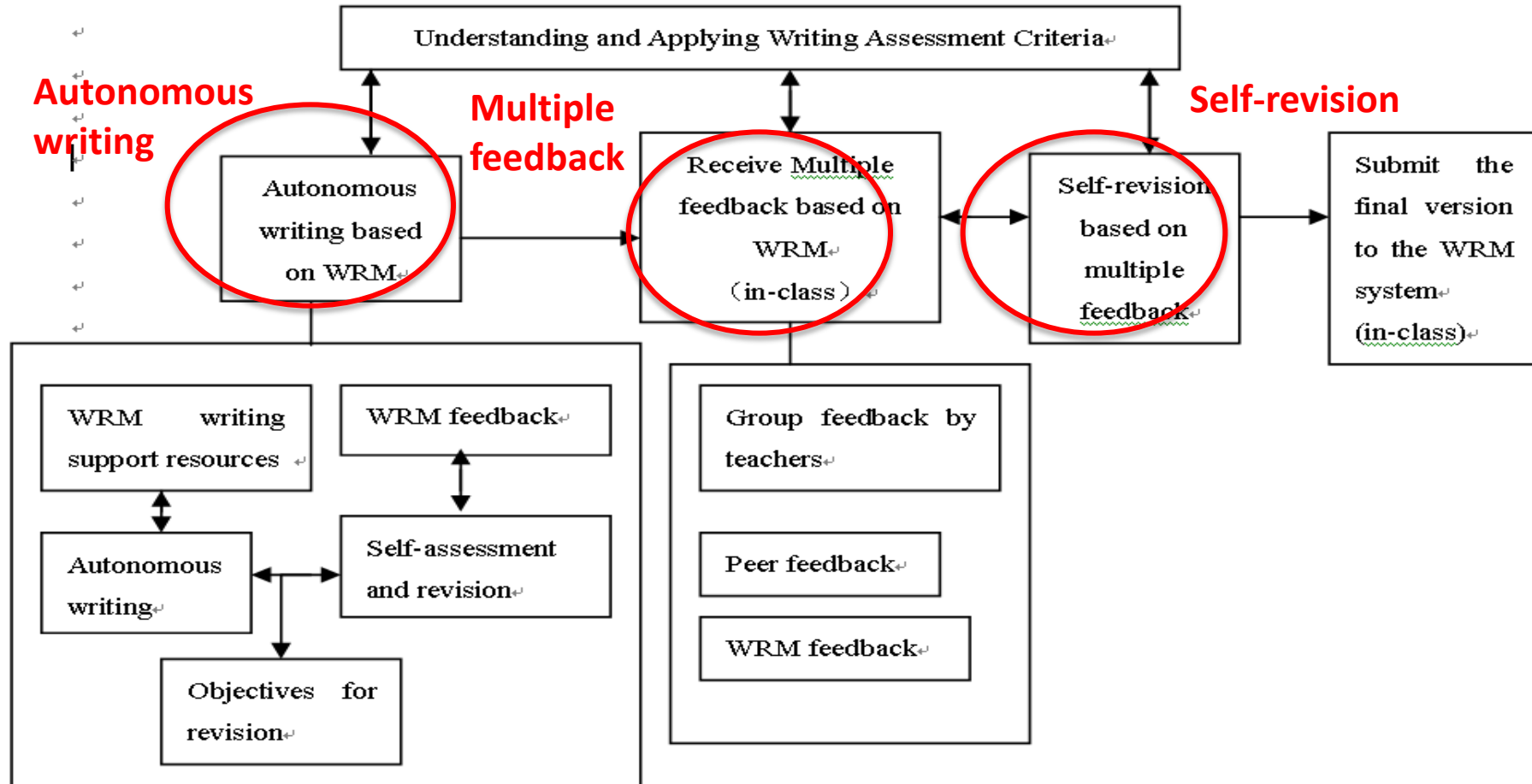
Writing rubrics



In addition to narrative feedback on their essays, students will receive a holistic score as well as analytic scores based on six writing dimensions: Ideas and Content, Organization, Voice, Word Choice, Fluency, and Conventions

Intervention 1: Working Model of Integrating WRM in Writing Instruction

Assessment criteria



Intervention 2: Teacher support

Content

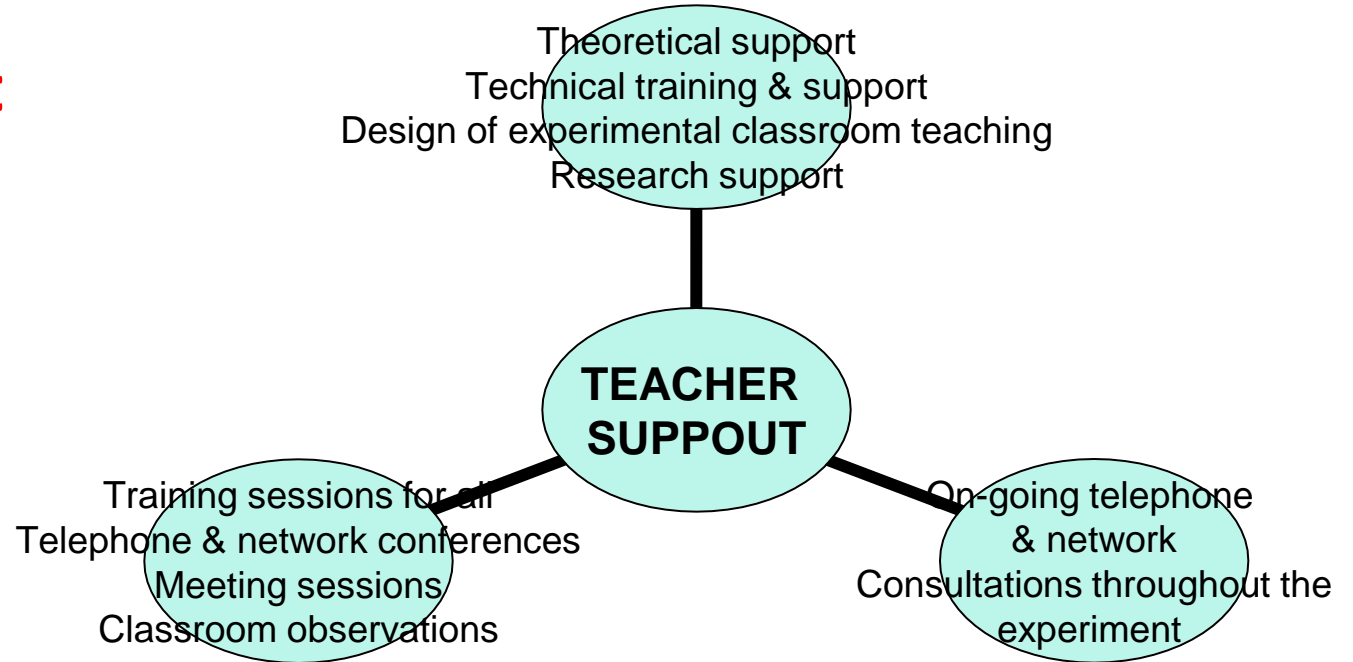
Theoretical support
Technical training & support
Design of experimental classroom teaching
Research support

**TEACHER
SUPPORT**

Form

Training sessions for all
Telephone & network conferences
Meeting sessions
Classroom observations

On-going telephone
& network
Consultations throughout the
experiment



Data Analysis:

- GLM analysis
- Content analysis

Research Results

1. Impact on the writing performance

Descriptive statistics

Variable	N	Pre-test mean	Pre-test SD	Post-test mean	Post-test SD	Mean post-pre test score gain	Score gain SD	Effect size
Experimental	224	2.34	0.57	2.86	0.74	0.52	0.85	0.79
Control	236	2.40	0.53	2.51	0.62	0.11	0.73	0.19
Experimental English major	37	2.65	0.61	3.56	0.64	0.91	1.05	1.46
Experimental non-English major	187	2.28	0.54	2.72	0.68	0.44	0.79	0.72
Control English major	33	2.70	0.64	2.90	0.67	0.20	1.09	0.31
Control non-English major	203	2.35	0.49	2.45	0.59	0.10	0.66	0.18

Gain score GLM analysis

Source	DF	SS	MS	F	Pr > F
Experimental or control group	1	18.75	18.75	31.46	0.0001
English major or non-English major	1	4.93	4.93	8.05	0.005
Group*English major	1	2.03	2.03	3.32	0.07
Experimental English major or non-English major	1	6.72	6.72	9.66	0.002

2. Impact on the writing process and learners

A. The integration of teacher, students and WRM assessment and feedback **enhanced interaction and motivated students to write and revise.**

- 70% students tend to write and revise more;
- 62.3% revise 1-2; 27.9% revise 3-4;
- I can remember more clearly when I correct my own mistakes through WRM.

(Source: University C, Student 1, Questionnaire)

B. Students learn to **use the writing assessment criteria** to guide their own writing.

The teaching experiment helped me to know better about the ideas and structure of English essays, it also helped to improve my self-assessment ability. Now I can see very clearly the strengths and weakness of an essay.

(University A, Student 1, Questionnaire)

C. Students became more autonomous via dynamic interaction with the AI-assisted writing evaluation tool and teacher feedback, correcting their mistakes and revising their essays.

What I found most attractive about the system was that it could force me to revise my essay, which improved my autonomy and writing.

(Source: University B, Student 2, Questionnaire)

3. Impact on the writing teaching process and teachers

A. A shift of focus From language form to content and discourse;
from product to process

*My attention used to be on the writing products, but now it shifts to the writing process, learners' affective change and self-regulating abilities.
(22-ZXQ, Post-Q)*

B. Attending to teaching/learning process

- (1) A pre-writing phase introduced with a focus on content
- (2) Interpretation of assessment criteria (AC) as an important part of teaching

In face of so many learners and so many problems in their writing, a teacher's efforts are but an ant's push, leaving him-/herself helpless. Helping learners grasp the AC means teaching them ways of knowing. As a result, learners are empowered and teachers liberated. ... It can help learners with their life-long writing. (10-TYY, Post-Q)

(3) More effective feedback offered

- By the AI-assisted writing evaluation tool: Immediate, locating the type of problem, competent in helping learners with language form
- By the teacher: Concrete, targeted and contextual

Feedback from the AI-assisted writing evaluation tool is relatively general. It can tell me roughly where my students are, with reference to native-speaker performance. My feedback is very concrete, related to the topic concerned and the context, with more concern for content and rhetoric. (08-WXY, Post-Q)

C. Change in teacher roles

Teacher as dominator, only assessor vs. as facilitator, co-assessor, senior learner, co-manager of learning, and researcher

Teachers now assume a variety of roles: facilitators in learning, organizers and coordinators of writing teaching/learning activities, trainers in technical skills and learning strategies, assessors to fill in the gaps left by the system in its feedback, analyzers of learner needs, senior learners concerning the AC and doubts about the system' feedback and scoring, researchers of their own teaching for the sake of improving teaching and self, ... (Mid-term report, XNSY)

Summary of major findings:

From the learner and the learning process:

- Improved performance
- Enhanced interaction
- Motivating
- Promoting autonomy

From the teacher and the teaching process:

- A shift of teaching focus
- Attending to the teaching/learning process
- Positive changes in teacher roles

3. Lessons

Writing instruction

1. AI-Assisted writing evaluation can serve as an effective teaching assistant by offering continuous writing support and dynamic assessment, thus enhances writing instruction.

Technology

2. The key to technology use might be neither hardware nor software, but rather human ware (Warschauer & Meskill, 2000).
3. Technology, when applied properly, can serve as a catalyst that leads to positive changes in teachers and their teaching.

Teachers

4. Yet, it is teachers, rather than technology, who determine the quality of teaching by intelligently exploiting the use of technology.
5. Teacher training and support, for example in the area of TPACK (Technological Pedagogical Content Knowledge, 整合技术的学科教学知识), are crucial to the introduction of any innovations in teaching.

Selected references:

Black, P. J., & William, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139–148.

Tang, Jin-lan & Rich, Changhua Sun (2017). Automated writing evaluation in an EFL setting: Lessons from China. *JALT CALL*, 13(2): 117-146.

Tang, Jin-lan & Wu, Yi'an (2017). Impact of AWE rubrics and automated assessment on EFL writing instruction. *International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT)*, 7(2): 58-74.

Warschauer, M., & Meskill, C. (2000). Technology and second language learning. In J. Rosenthal (Ed.), *Handbook of undergraduate second language education* (pp. 303–318). Mahwah, New Jersey.

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

Wu, Yi-an & Tang, Jin-lan (2012). Impact of integrating an automated assessment tool into English writing on university teachers *Computer-Assisted Foreign Language Education* 146 (4): 3-10.



北京外国语大学
Beijing Foreign Studies University

人工智能与人类语言
Artificial Intelligence and Human Languages Lab
重点实验室



北京外国语大学
网络教育学院
www.beiwaionline.com

Thank you !

tangjinlan@bfsu.edu.cn,
tangjinlan@beiwaionline.com



Online Institute of Education, BFSU



AI and Human Languages Lab, BFSU

LANGUAGE LEARNING IN THE DIGITAL AGE / 数字时代的语言学习
法国驻华大使馆文化教育合作处、德国歌德学院联合主办
2020年9月4日-6日