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
Cradle of Diplomats  外交官的摇篮。

Wherever a five-starred red flag is flying, there are Beiwei 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。
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](http://www.chinacall.org.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 & Wiliam, 1998, p. 140
AI-assisted writing evaluation

• Al-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）
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
Intervention 2: Teacher support

Content

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

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Pre-test mean</th>
<th>Pre-test SD</th>
<th>Post-test mean</th>
<th>Post-test SD</th>
<th>Mean post-pre test score gain</th>
<th>Score gain SD</th>
<th>Effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>224</td>
<td>2.34</td>
<td>0.57</td>
<td>2.86</td>
<td>0.74</td>
<td>0.52</td>
<td>0.85</td>
<td>0.79</td>
</tr>
<tr>
<td>Control</td>
<td>236</td>
<td>2.40</td>
<td>0.53</td>
<td>2.51</td>
<td>0.62</td>
<td>0.11</td>
<td>0.73</td>
<td>0.19</td>
</tr>
<tr>
<td>Experimental English major</td>
<td>37</td>
<td>2.65</td>
<td>0.61</td>
<td>3.56</td>
<td>0.64</td>
<td>0.91</td>
<td>1.05</td>
<td>1.46</td>
</tr>
<tr>
<td>Experimental non-English major</td>
<td>187</td>
<td>2.28</td>
<td>0.54</td>
<td>2.72</td>
<td>0.68</td>
<td>0.44</td>
<td>0.79</td>
<td>0.72</td>
</tr>
<tr>
<td>Control English major</td>
<td>33</td>
<td>2.70</td>
<td>0.64</td>
<td>2.90</td>
<td>0.67</td>
<td>0.20</td>
<td>1.09</td>
<td>0.31</td>
</tr>
<tr>
<td>Control non-English major</td>
<td>203</td>
<td>2.35</td>
<td>0.49</td>
<td>2.45</td>
<td>0.59</td>
<td>0.10</td>
<td>0.66</td>
<td>0.18</td>
</tr>
<tr>
<td>Source</td>
<td>DF</td>
<td>SS</td>
<td>MS</td>
<td>F</td>
<td>Pr &gt; F</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------------------------------</td>
<td>----</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>--------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental or control group</td>
<td>1</td>
<td>18.75</td>
<td>18.75</td>
<td>31.46</td>
<td>0.0001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>English major or non-English major</td>
<td>1</td>
<td>4.93</td>
<td>4.93</td>
<td>8.05</td>
<td>0.005</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group*English major</td>
<td>1</td>
<td>2.03</td>
<td>2.03</td>
<td>3.32</td>
<td>0.07</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental English major or non-English major</td>
<td>1</td>
<td>6.72</td>
<td>6.72</td>
<td>9.66</td>
<td>0.002</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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:


Thank you!

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

Online Institute of Education, BFSU

AI and Human Languages Lab, BFSU