Advances in Educational Technology and Psychology (2024)
Clausius Scientific Press, Canada

DOI: 10.23977/aetp.2024.080623
ISSN 2371-9400 Vol. 8 Num. 6

A Critical Review of Research on ChatGPT-assisted L2 Education: A Formative Assessment Perspective

Dong Yao[1,a,*], Qi Zhang[1,b]

[1]*Department of English, College of Humanities and Foreign Languages, China Jiliang University, Hangzhou, Zhejiang Province, China*
[a]*donnieyao@cjlu.edu.cn*, [b]*ChanthyZhang_25@163.com*
[*]*Corresponding author*

Keywords: Review, ChatGPT, second language education, formative assessment

Abstract: With the advent and advancement of ChatGPT, both theoretical and empirical research over its impact on second language education has become generally prevalent. The current paper reviews the empirical research on ChatGPT-assisted second language education to highlight the essence of process-centric reform for education corresponding to the formative assessment from the perspectives of labyrinthine stakeholders (i.e., students, teachers and peers). However, the results showed that the majority does not pertain to peer assessment, psychological inclination together with practical application for the upgrading of teacher-assessment. Besides, some studies published in Chinese propose corresponding framework yet lack direct data support, thus lying down at the superficial level. The results from the current paper suggest that further research could focus more on the interaction between learner- and peer- rapports with the introduction of ChatGPT and conduct multifarious valid studies for assessment reform.

1. Introduction

The past decades have witnessed a surge in interest in English-language programs within higher education, driven by the demand for accessible second language learning resources. This trend has prompted the development of AI-mediated learning tools such as Grammarly and Google Translate, which have revolutionized teaching and assessment methods to align with students' evolving needs [5,20,21]. Kooli pointed out the transformation effect of young students in learning and interaction with technology while ChatGPT[11], in particular, has emerged as a versatile tool for various communication purposes, collaborative data analysis, and peer-review processes, which align closely with the principles of formative assessment, which prioritize ongoing feedback and student engagement. Given the absence of systematic vindication studies for second language with the succor of ChatGPT, a review paper is warranted to agglomerate the majority of present theoretical papers for ensuring the education reform and dwindling potential risk of indulging learners. From there, the current paper critically reviews the present appropriate research on ChatGPT-empowered educational exercise and L2 education under the structure of formative language assessment, focusing on multiple stakeholders (i.e., students, teachers, and peers) of assessment. The results of the paper may bring about revelatory implications to the field of L2 teaching and learning.

2. Formative Assessment Framework

Formative assessment in higher education, particularly in the context of English teaching, has garnered significant attention in recent literature. Its purpose is to provide feedback to teachers and students during the course of learning about the gap between students' current and desired performance so that action can be taken to close the gap[1]. Unlike summative assessments to evaluate students at the end of a unit or course, formative assessments are designed to be embedded in classroom learning processes continuously[10,13]. Formative assessment consists of three types of stakeholders, namely teachers, peers and students themselves (see Figure 1). [22] argues that both peer and self-assessment can be important vehicles for providing feedback on students' existing performance and steps for moving forward[22], especially for self-referenced assessment, which can reduce fierce competition[16] and improve learner outcomes, increase confidence, increase independence, change the culture of classroom and leverage their nature of learning for curricular and instructional improvements.

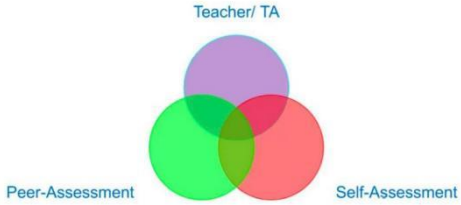

Figure 1: Subjects of Formative Assessment

2.1 Students

Formative assessment empowers students with more internal motivation and initiative. Primarily, research has shown that formative assessments can have a positive impact on student motivation for learning[2]. Furthermore, Nicol and Macfarlane-Dick further recommend discussions with the instructor and with classmates that may be highly motivational and may also prompt students to view their own work with greater detachment[15]. One study by Yeh revealed that students who received feedback completed more work with greater accuracy than students who did not receive feedback[29]. Brown et al. focused on student self-assessment as a key component of formative assessment[3]. They highlight the importance of understanding accuracy in self-assessment to enhance academic achievement and self-regulatory learning, which ChatGPT contributes to.

2.2 Teachers

Teachers are primarily expected to elicit evidence of achievement and use feedback to modify their teaching and learning activities. Accurate for feedback and assessment, formative assessment provides educators with strong accuracy by the complexity of assessments, akin to oral presentations, reports and essays[12].

2.3 Peers

Peers are analogous to teachers in that they have unique insights into learning and their relational power will facilitate others to be open to instructional strategies. As a part of formative assessment, standard-referenced reporting allows students to stay with their cohort groups—their grade-level peers[12]. In addition, peer review and feedback provided an authentic context for evaluation and monitoring of works-in-progress[27]. Formative assessment brings a collaborative and reflective

learning environment that gives students a deeper understanding. It is recommended that education should encourage peer and teacher dialogue rather than viewing learning as a process that involves a unidirectional transmission of information[16].

Based on the above-mentioned review, it is not difficult to find that the present educational reform has shifted its perspective over acquisition process and student-centrism since Reinforcement Learning from Human Feedback (RL-HF) has been appended into ChatGPT-3[23], which is interrelated to multifarious formative assessment, breaking report or essay boundaries and purifying the essence of education[31].

3. Selection of Studies

In request to collect relevant and methodologically sound research, the literature review was intended to identify published studies investigating the impact of language assessment in L2 classes and general education. The analysis was carried out by examining the most widely-used electronic academic databases: i.e., Web of Science, Elsevier, Springer, and CNKI. The keywords, akin to ChatGPT, education, and further L2, were the predominant retrieving references. Since ChatGPT, an AI-powered chatbot, was first introduced to the public in November 2022, the time limit has been set from 2022 to 2024. These studies were further screened through the following dimensions: (1) studies mentioning ChatGPT-assisted approaches distinctively; (2) studies designing for higher education; and (3) studies concentrating on procedural feedback as formative assessment. The abstracts and full texts of the remaining 15 papers were evaluated and all were eventually included due to their relevance to this review paper.

3.1 Existing Studies

Table 1: Empirical Studies on ChatGPT-aided General and L2 Education

Author(s) and year of study	Focus/ Foci of research	FA inferences
General education		
[26]	ethical risks and avoidance approaches	TA
[30]	educational transformation	TA; SA
[7]	academic writing, accuracy	TA
[14]	automated essay scoring (AES)	TA; SA
[24]	activity-based learning	TA
[33]	cultural-awareness; biases	PA
[9]	teacher-and ChatGPT relationship	TA
[23]	internal mechanism; compound brains	TA; SA
L2 Education		
[6]	catalyst (emotion)	SA
[4]	L2 writing	TA
[31]	L2 personalized schemes	TA; SA
[32]	L2 vocabulary acquisition	SA
[28]	efficiency; plagiarism	SA
[19]	English teaching reform	SA
[8]	English teaching reform	SA

Note. FA=Formative assessment; TA=Teacher assessment; SA=Student assessment; PA=Peer assessment

There are 15 studies on language assessment for AI-assisted higher education (see Table 1). Among all studies, seven out of which have been conducted on L2 classes, and the left on general

education classified by multiple stakeholders. The foci of these studies lie in the cultivation of professional skills, but they all follow the formative assessment with strong process-centered inclination. What follows is the critical review of these studies and the findings of which will be interpreted in the context of the formative assessment dimensions.

3.2 Students

From the above table, it is found that nine studies[6, 8, 14, 19, 23, 28, 30, 31, 32] are linked to the student-assessment inference.

General Education: As for general education, through engaging discussions, [27] noted that ChatGPT is a versatile AI language model in aiding students[27]. It can serve as a virtual one-to-one tutor, providing explanations and assistance in subjects, act as a language practice partner for language learning, offer writing support and feedback, act as a stress-relief conversational partner, boost curiosity and exploration of new topics, help in exam preparation, aid in personal organization and time management, and promote critical thinking.

As for learners themselves, ChatGPT can function as a catalyst to render local classroom needs and provide a list of feasible solutions for diverse educational needs at the instructional and assessment levels, i.e., a stimulus for teachers and students. [32] name the benefits of ChatGPT support, akin to personalization, exploration and diversity following students' preferences for time, place and pace[31]. In this respect, ChatGPT, GPT-4 with its powerful architectural modeling technology and features realize the content generation and multi-round dialog level, arouse students' interests for independent study and strengthen their predominance[19]. What garnered substantial attention are the improvable shortcomings in accuracy, data pollution, ethnics and plagiarism whereas some scholars consider them beneficial chances for students to form critical thinking even though the volume of ideas unmatched to ChatGPT generalization[23,30].

L2 Education: With enough engagement, scholars hold the belief that dialogic acquisition process satisfies students to perform self-directed learning and wholeheartedly involve in. ChatGPT assisted them to brainstorm, automatically examine learners' language proficiency and transfer their traditional role for close update of learning[31]. Compared with the rote or repetitive memorization for vocabulary acquisition, Zhang's research illuminated the reception and production empowered by repetitive dialogs and stimulus, which somehow transformed the passively received pattern. On top of that, Ghafouri indicated that ChatGPT is beneficial for creating a teacher-and learner-friendly environment and increasing the learners' grit through psycho-emotional network[6].

Even though ChatGPT brings about unprecedented blessings, the educational field is still instilled with concerns. Yan investigated the empirical study for L2 class through one-week and small-scale practice, finding that the L2 writing learners are generally perturbing instead of jubilant[28]. Integrity, capacity for critical thinking and cross-cultural awareness[24][31] handicapped both students and teachers to give ChatGPT a full rein. But the fear over assessment equality and fairness mirrors the entrenched scare of unexpected consequences and assessment result; instead, enough wiggled room and mistakes enrich users' experiences for this revolution. [8] proposed that student-learning are shifting from the fragmented, traversal self-thinking processing state to the prefabricated and borrowed machine processing state, which by contrast rebelled the truth. If students take the latter one, they can never form critical thinking but to add nuanced elements to the prefabrication[8]; instead, it is those fragmented information that keeps stimulating their brain and ingraining knowledge into their minds.

These studies touch upon the credits sides and convey concerns over-dependence, unfairness and unreliability for ChatGPT. However, according to the questionnaires handed out by [25], one third of respondents declared their willingness to use it in the future out of expectations[25]. Research over

different levels of new-comers remains unknown. Nonetheless, all above didn't touch upon the essence of over-reliance and substitute concerns.

3.3 Teachers

In terms of teacher-assessment inference, eight studies [4, 7, 14, 23, 24, 26, 30, 31] are listed.

General Education: [30] discussed how ChatGPT's sophisticated text generation can transform classroom dynamics, offering teachers new ways to enhance instruction and administrative functions[30]. A study by [9] found that 9 out of 11 teachers use ChatGPT chat history to assess subconscious and relaxed acquisition, shifting from mark-orientation to progress-orientation. However, the effectiveness of this approach and its impact on students' subconscious acquisition are unknown, raising concerns about man-machine ethics. Teachers must design personalized schemes and monitor history to cultivate critical and ethical thinking[30,7], adding pressure as they must assess prompt quality. Long-term feasibility should be investigated if efforts do not outperform traditional methods.

Generally speaking, [23]'s research demystified the internal mechanism of self-adaptation approach. Trainers train ChatGPT to form rules. Correspondingly, teachers show students tasks, which function as pre-rules. During the dialogues, students uplift their efficiency through repetitive trails, errors and questions. It is crucial that students can provide high-quality answers on the ground of brainstorming and healthy mindsets.

L2 Education: As for ChatGPT-assistance for L2 writing, [17] supplemented retrospection[17] and laid foundation for discussing the activities for preparation and organization[4]. Grounded on the traditional PWP writing class pattern, ChatGPT can be a counsellor for logical problems apart from basic language mistakes instantaneously. In the long run, teachers can slightly get away from responsibility for teaching design, cultural output for targeted language but call for the examination over prompts. [4] also elucidated the significance of the reflection stage for expanding the zone of proximal development to stimulate students[4].

Teachers can compare references with students to distinguish text accuracy and felicity, spurring writing interests and integrating pedagogical and professional knowledge. [31] empirically examined ChatGPT's role in L2 vocabulary acquisition, emphasizing its efficacy in an innovative interactive learning environment and providing teaching strategies to enhance cooperation[31]. [26] proposed ethical risks and avoidance methods for educational ChatGPT to strengthen teacher-student rapport[26]. [24] addressed integrity issues as ChatGPT comments increasingly resemble human-written feedback[24]. [14] showed AES's accuracy and reliability, highlighting human evaluation issues like fatigue, subjectivity, and inconsistency, especially for beginners[14]. Automatic grading and feedback tools (AGFTM) contribute to reducing the burden of student performance assessment, requiring further datasets to allow teachers to adjust strategies[1]. [30] proposed transformations for teacher development, including embracing AI, lifting digital competence, collaborating with AI, advancing technological pedagogy, enhancing student-centric learning, nurturing responsible AI engagement, etc[30]. Researchers noted diverse teacher-party assessment preparations but overlooked teachers' active engagement and pre-training efforts.

3.4 Peers

Merely two out of 15 studies authentically mention the peer-assessment inference. As for the accuracy of peer assessment, [33] analyzed the WEIRD (western, educated, industrial, rich and democratic) cultural values embedded in moral dilemma stories generated by ChatGPT 3.5[33]. However, as the authors themselves point out, morality is time- and culture-specific, and trainers are replete with outdated biases and stereotypes that can present erroneous feedback and confuse

stakeholders for assessment. In such a jumbled environment of values, students are deeply influenced by their corresponding peers.

Besides, the study held by [28] claimed that students assimilated enormous advantages from ChatGPT to outperform their peers. In the long run, students themselves will lift the bar, and provoke competitions. Nonetheless, formative assessment alleviates the fierce competition within intelligence education and endows them with opportunities for moral progress. To recap, ChatGPT-aided classes provide more room for all-round development if supported by a felicitous learning environment.

4. Conclusion

The current paper reviewed 15 studies related to the feasibility of ChatGPT-assisted L2 general education based on the formative assessment. First, it was found that most studies focused on educational efficiency, students' motivation, with each researcher distributing several strokes for other universal concerns, plagiarism, ethnics and to name a few. Second, most studies accentuate the significance of teachers and students, while for peers, they barely mention thus putting students under a solo study environment. In addition, peer assessment catalyzes applicable learning environments such that peer influence also merits academic research attention. At present, ChatGPT is clearly, by and large, an irreducible member, which appeals to future bursting validation studies ought to be put on agenda. And, more attention should be paid to low-quality beginners with uneven language proficiency regarding their attitudes towards ChatGPT and users who entertain strong disinclination. Finally, the most researches remain theoretical and took granted anachronistic theories into consideration without solid foundation, then presented theoretical conclusions as if students were in the sterile laboratories.

In conclusion, this review paper highlights future research directions to gain more theoretical and practical insights into the application of ChatGPT to L2 learning and the corresponding pedagogical adjustments. Given the complexity and deterioration of initiatives within formative assessment in the new era, reformers should highlight self-assessment and peer assessment as ways to save teachers unnecessary effort and alleviate the increasingly pressure led by *Double Reduction* policies. Further research ought to shed light on data collection across demographics and contexts, such as data augmentation, crowd-sourcing and domain adaptation[18] with the purpose of ensuring its accuracy to large extent.

Acknowledgment

Funding Information: This paper is funded by Zhejiang Federation of Humanities and Social Sciences Circles (No: 24NDQN206YBM), and the project title is: *Investigating the effect of ChatGPT on L2 education*.

References

[1] Assessment, F.T., Heritage, M. (2008) Learning Progressions: Supporting Instruction and Formative Assessment. The Council of Chief State School Officers.

[2] Banić, B., Konecki, M., Konecki, M. (2023). Pair programming education aided by ChatGPT. The 46th MIPRO ICT and Electronics Convention (MIPRO).

[3] Brown, G.T., Andrade, H.L., Chen, F. (2015). Accuracy in student self-assessment: directions and cautions for research. Ass in Edu. 22(4), 444-457.

[4] Chen, M., Lyu, M.(2024). College English Writing Teaching in the Context of ChatGPT. Con For Lang Stu. 1, 161-168.

[5] Del Gobbo, E., Guarino, A., Cafarelli, B., Grilli, L., Limone, P. (2023) Automatic evaluation of open-ended questions for online learning. A systematic mapping. Stu in Edu Eva. 77, 101258.

[6] Ghafouri, M. (2024). ChatGPT: The catalyst for teacher-student rapport and grit development in L2 class. Sys. 120, 103209. https://doi.org/10.1016/j.system.2023.103209.
[7] Halaweh, M. (2023). ChatGPT in education: Strategies for responsible implementation. Con Edu Tech. 15(2), ep421.
[8] Hu, J., Qi, Y. (2023). Chinese Foreign Language Education in the Era of ChatGPT: Seeking Change and Responding to Change. For Lan Com Tea. 1, 3-6.
[9] Jeon, J., Lee, S. (2023). Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT. Edu Inf Tech. 28, 15873–15892.
[10] Kariri, K. A., Cobern, W. W., Al Sultan, A. A. (2022). Investigating high school science teachers' readiness for implementing formative assessment practices. Eurasia Journal of Mathematics, Sci and Tech Edu. 18(12), em2188.
[11] Kooli, C. (2023). Chatbots in Education and Research: A Critical Examination of ethical implications and solutions. 15(7), 5614.
[12] Marzano, R. J. (2010). Formative assessment & standards-based grading: Classroom Strategies that Work. Marzano Research.
[13] Menández, I. Y. C. et al. (2019). The importance of formative assessment in the learning teaching process. Int Jour of Soc Sci and Hums. 3(2), 238-249.
[14] Mizumoto, A., Eguchi, M. (2023). Exploring the potential of using an AI language model for automated essay scoring. Res Meth in App Lin. 2(2), 100050.
[15] Nicol, D. J., Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. Stu in High Edu. 31(2), 199-218.
[16] Nitko, A.J., Brookhart, S.M. (2006). Educational Assessment of Students (5th Edition). Prentice-Hall, Inc.
[17] Liu, L. et al. (2023). ChatGPT Arrives: New Opportunities and Challenges for International Chinese Education (Part 1). Lan Tea and Lin Stu. 3, 1-14.
[18] Lo, C. K. (2023) What is the impact of CHATGPT on education? A rapid review of the literature. Edu Sci. 13(4), 410.
[19] Lu, Y. (2023). Exploring the Impact of Artificial Intelligence on English Teaching and Its Reform Path. Open Jou of Poly Uni. 3, 54-58.
[20] 2023 Education in the Post COVID-19 Era -- Opportunities and challenges. Lecture notes in educational technology.
[21] Ou, A. W., Stöhr, C., Malmström, H. (2024). Academic communication with AI-powered language tools in higher education: From a post-humanist perspective. Sys. 121, 103225.
[22] Sadler, D. R. (1989). Formative assessment and the design of instructional systems. Ins Sci. 18(2), 119-144.
[23] Shen, S., Zhu, Z. (2023). ChatGPT-like Products: Internal Mechanisms and Their Impact on Learning Evaluation. Chi Dis Edu. 4, 8-15.
[24] Shloul, T. A. et al. (2024). Role of activity-based learning and ChatGPT on students' performance in education. Computers and Education. Art Int. 6, 100219.
[25] Ubowska, A., Królikowski, T. (2023). ChatGPT - opportunities or threats in the educational process. Pro Com Sci. 225, 4551-4559.
[26] Wang, Y., Wang, D., Liang, J., Liu, C. (2023). Ethical Risks and Avoidance Strategies of ChatGPT's Educational Application. Open Edu Res. 2, 26-35.
[27] Xiao, Y., Yang, M. (2019). Formative assessment and self-regulated learning: How formative assessment supports students' self-regulation in English language learning. Sys. 81, 39-49.
[28] Yan, D. (2023). Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation. Edu and Info Tech. 28(11), 13943-13967.
[29] Yeh, S. S. (2008). The cost-effectiveness of comprehensive school reform and rapid assessment. Edu PolAna Arch.16: 13.
[30] Yu, H. (2024) The application and challenges of ChatGPT in educational transformation: New demands for teachers' roles. Heli. 10(2), e24289.
[31] Zhang, Z., Hong, H. (2023). Foreign Language Teaching Supported by ChatGPT: Empowerment, Issues, and Strategies. For Lang Wo. 2, 38-44.
[32] Zhang, Z., Huang, X. (2024). The impact of chatbots based on large language models on second language vocabulary acquisition. Heli. 10(3), e25370.
[33] Zheng, Y., Stewart, N. (2024). Improving EFL students' cultural awareness: Reframing moral dilemmatic stories with ChatGPT. Computers and Education. Art Int. 6, 100223.