OPEN ACCESS

International Journal of Language Learning and Teaching Research
IJLLTR
Volume 1 (2026)
https://doi.org/10.65553/IJLLTR.260105

Computerized Dynamic Assessment and Its Effects on Second Language Production: A Meta-Analytic Perspective

Han Shi *Xiaolei Ma*
Basic Education College, National University of Defense Technology, 410000, China
Corresponding author: maxiaolei@nudt.edu.cn

Abstract

A substantial body of research in the field of second language acquisition has demonstrated that computerized dynamic assessment (C-DA) has positive effects on learners' second language (L2) proficiency. However, a comprehensive research synthesis concerning the overall effects of C-DA on L2 production remains limited. To address this gap, the present study employs meta-analytic methods and incorporates data from L2 studies with 889 participants to quantify its effects on L2 production. The main findings are as follows: (1) C-DA enhances learners' writing and speaking abilities, with a particularly significant positive effect on oral production; (2) C-DA has more pronounced effects on the fluency and accuracy of both written and oral output; and (3) although intervention duration, learners' L2 proficiency, and educational level do not significantly moderate the overall effects, all subgroups show certain positive intervention effects, with more pronounced effects observed for short-term interventions, intermediate-to-upper-intermediate learners, and university learners. Drawing on the synthesis of empirical research results, this study

Submission date: May 1, 2026; Revision date: November 1, 2026; Online: November 12, 2026

Copyright: © 2026 The author(s). This is an open access article distributed under the terms of the Creative Commons Attribution Non-Commercial 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

intends to enrich the understanding of C-DA and L2 acquisition, eventually offering implications for L2 writing and speaking instruction.

Keywords: computerized dynamic assessment; second language production; L2 writing; L2 speaking; meta-analysis

Introduction

Dynamic Assessment (DA) originates from the concept of the Zone of Proximal Development (ZPD) in Vygotsky's (1978) sociocultural theory. Unlike traditional assessment approaches that primarily focus on learners' current performance, DA underscores the mediated interaction among the assessor, the learner, and the assessment task, integrating personalized intervention with assessment (Feuerstein et al., 1988). It intends to overcome the limitations of the separation between teaching and assessment, attempting to stimulate the learners' potential. With the rapid development of information technology, the affordance of digital technology in second language (L2) education has profoundly reshaped traditional pedagogical practices. Therefore, integrating digital technologies with dynamic assessment has become an inevitable trend, leading to the emergence of Computerized Dynamic Assessment (C-DA). C-DA refers to a form of immediate dynamic assessment conducted through a computer platform, which employs algorithm-based models to produce immediate, dynamic, and individualized diagnostic feedback (Poehner & Lantolf, 2013). Furthermore, the defining feature of C-DA lies in the capacity to provide continuous mediation that responds to learners' difficulties and is aligned with their ZPD. Through analyzing learners' real-time online performance, C-DA flexibly adjusts the intervention strategies and enhances the standardization of mediation thereby probing L2 learner's potential to surpass existing abilities (Lu et al., 2026).

While research on C-DA in L2 learning has expanded rapidly, many studies predominantly focus on its effectiveness in fostering learners' language proficiency. Early studies examined not only broad language skills, including listening, speaking, reading, writing, translation, and grammar (Mardian & Nafissi, 2022; Meng et al., 2024; Pileh Roud et al., 2021; Yang & Qian, 2020), but also more specific areas, such as paraphrasing competence and the comprehension of conversational implicatures (Geng et al., 2022; Qian et al., 2023; Qin & van Compernolle, 2021). Recent studies have shifted their focus toward examining the intervention effects of different forms of dynamic assessment, particularly comparisons between C-DA and interactive dynamic assessment approaches, such as group and rater mediated dynamic assessment (Ghahderijani et al., 2021; Sherkuziyeva et al., 2023). More importantly, a review of the C-DA literature reveals numerous empirical investigations targeting productive skills, particularly writing and speaking, with substantial variation among them (Ebadi & Vakili, 2019; Jiang, 2026; Li et al., 2020; Sherkuziyeva et al., 2023; Xiao, 2024). Meta-analysis, a quantitative research methodology, enables researchers to synthesize findings from multiple empirical

studies, calculate overall effect sizes, and identify potential moderators influencing research outcomes (Cai, 2012). Given that an existing meta-analysis has already established the overall positive effects of C-DA on broad measures of L2 performance (Lu et al., 2026), the present study specifically explores its effects on the written and oral production of English as a foreign language (EFL) learners. By adopting a meta-analytic approach and integrating effect sizes from previous studies, this study is expected to yield empirical evidence that inspire educators in designing applications for C-DA and even artificial intelligence-enhanced dynamic assessment in L2 education.

Literature Review

C-DA can be defined as an immediate assessment approach that relies on digital platforms to evaluate learners' current performance and provide standardized mediation (Geng et al., 2022; Tai & Wang, 2022). C-DA was subsequently developed as a technology-mediated extension of DA (Lu et al., 2026). DA is fundamentally grounded in Vygotsky's (1978) sociocultural theory (SCT). The notion of ZPD under the framework of SCT insists that learners' emerging abilities can be identified through their performance under appropriate assistance, combining educational psychology theories and advanced information technology. C-DA delivers mediation through computerized procedures (Poehner & Lantolf, 2013). It employs computer-supported adaptive feedback and analytical techniques to overcome the limitations of traditional static assessments, ultimately aiming to predict and promote learners' potential development. Feuerstein et al.'s (1988) concept of mediated learning experience (MLE) further highlights the role of mediation in cognitive development. From the perspective of C-DA, learner development depends not only on exposure to a task but also on sufficient assistance. In a computerized environment, the digital platform is expected to perform some functions conventionally undertaken by a teacher or assessor, such as diagnosing errors promptly and offering adaptive intervention. Nevertheless, computer delivery alone does not guarantee meaningful mediation. Through such mediation, learners strive to overcome individual cognitive constraints and improve their proficiency.

Dynamic assessment interventions can generally be categorized into two major types: interactionist and interventionist DA (Lantolf & Poehner, 2004). Existing research has primarily adopted the interventionist approach, because pre-designed mediation is easier to standardize and deliver to large groups of learners efficiently. Based on when mediation is offered, Sternberg and Grigorenko (2002) proposed two general DA formats: the "cake" format and the "sandwich" format. The sandwich format consists of three sequential stages: pre-test, intervention, and post-test, requiring strict experimental control and systematic intervention design. In the cake format, mediation is embedded within the assessment sequence. This model was proposed by Guthke and Beckmann (2000), which intended to simulate children's language development with computer assistance. It advocates providing graduated prompts according to learners' ongoing performance, marking an initial attempt of technology-enhanced dynamic assessment.

Research on C-DA has gradually shifted from cognitive and educational psychology to second language acquisition. Correspondingly, C-DA research has progressed from theoretical exploration to practical applications, with increasingly refined research objectives, more targeted mediation procedures, and diversified

outcome analyses. The majority of C-DA studies have focused on discrete linguistic items (e.g., specific grammatical structures) or broad language skills (e.g., listening and reading) (Jiang, 2026). Existing studies have examined language proficiency development in specific linguistic dimensions, including paraphrasing, grammatical accuracy, and discourse coherence (Mardian & Nafissi, 2022; Qian et al., 2023; Yang & Qian, 2020). Moreover, researchers have conducted comparative studies by examining C-DA alongside other forms of dynamic assessment (e.g. group dynamic assessment), highlighting the relative advantages of standardized computer mediation and flexible interpersonal mediation. (Feng & Zou, 2026; Ren, 2019; Sherkuziyeva et al., 2023). Furthermore, researchers have increasingly incorporated multidimensional indicators to provide stronger evidence for the effectiveness of C-DA, among which complexity, accuracy, and fluency (CAF) measures and cognitive load assessments are particularly representative (Babamoradi et al., 2018; Ghahderijani et al., 2021; Imsa-ard & Barrot, 2026; Jiang, 2026).

Research on C-DA in L2 writing and speaking tasks has been relatively more extensive (Feng & Zou, 2026; Ghahderijani et al., 2021; Jalambo et al., 2025; Li et al., 2020). In terms of L2 writing, researchers have mainly focused on investigating the effects of C-DA on EFL writing proficiency and different forms of feedback provision (Xu, 2022; Yang & Qian, 2020). Li (2015) conducted a mixed-methods study combining pre-test, post-test, and control group designs to compare different English writing feedback approaches based on dynamic assessment theory. He found that computer-assisted monitoring and individualized feedback supported improvements in L2 writing. This aligns with recent empirical evidence demonstrating that tailoring written feedback to fit learners' individual characteristics, such as cognitive learning styles, significantly enhances EFL writing outcomes (Lu & Pan, 2026). With the advancement of research, scholars have extended the application of C-DA to more specialized writing contexts, including academic writing (Ebadi & Vakili, 2019) and email requests (Nicholas et al., 2026). Comparing teacher-student face-to-face dynamic assessment with C-DA, Ebadi and Vakili (2019) concluded that C-DA not only facilitated the identification of learners' developmental errors but also encouraged autonomous correction. These preceding findings suggest that the intervention effectiveness of C-DA can be extended to various functional text types. The emergence of generative artificial intelligence (AI) has created new possibilities for more adaptive forms of computer mediation. Shen et al. (2026) creatively compared chatbot-assisted dynamic assessment (CA-DA) and chatbot-assisted non-dynamic assessment (CA-NDA). This study emphasizes that the CA-DA group achieved comparable writing performance to both the CA-NDA and the comparison group. Nevertheless, it should be highlighted that the CA-DA approach yielded additional benefits in helping learners generate their ideas, developing their "maturing" ideas, and emulating an authentic communication environment. Hence, the contribution of CA-DA may also become visible in developing learners' writing ability.

Meanwhile, the application of C-DA in L2 speaking instruction has also become a prominent research focus in this field. The real-time corrective feedback mechanism of C-DA can improve the efficiency of traditional speaking instruction by rapidly and accurately identifying pronunciation deviations, grammatical errors, and pragmatic inadequacies in learners' oral production. C-DA generates individualized assessments and adaptive training programs tailored to learners'

specific needs. It not only reduces learners' anxiety during oral communication but also provides targeted support for enhancing their oral comprehension and production abilities (Ghahderijani et al., 2021). Sherkuziyeva et al. (2023) conducted a strictly designed control experiment comparing C-DA with traditional instructional assessment approaches to track the distinguished improvements of learners' speaking complexity, accuracy, and fluency. The findings demonstrated that learners receiving C-DA intervention achieved significant improvements in speaking fluency and logical coherence, further highlighting the advantages of the well-designed application of technology-enhanced dynamic assessment.

Existing research has consistently demonstrated that C-DA contributes to influencing learners' L2 proficiency. However, findings remain inconsistent and fragmented in individual studies due to heterogeneity in mediation design, assessment-related factors, learner populations, and study quality. For instance, Ren (2019) found that C-DA significantly improved learners' writing performance in terms of lexical density, lexical sophistication, and accuracy, whereas no significant effects were observed for lexical diversity, syntactic complexity, or fluency. In contrast, Sherkuziyeva et al. (2023) reported significant differences between the experimental and control groups in terms of post-test complexity, with the experimental group outperforming the control group in L2 production complexity. To synthesize existing empirical evidence, the present study introduces meta-analysis as a statistical approach. Compared with traditional narrative reviews, meta-analysis represents a quantitative secondary analysis of complicated research findings (Shen & Zhang, 2020). Meta-analysis enables researchers to systematically account for sampling errors and statistically non-significant findings (Plonsky & Oswald, 2015), quantify heterogeneity across studies, and identify potential sources contributing to variations in effect sizes. In recent years, this methodology has been widely adopted in the field of second language acquisition (Cai, 2012; Shen et al., 2025; Wang et al., 2024). Lu et al. (2026) synthesized 35 studies using a three-level meta-analytic model and reported large positive effects for both cake-format and sandwich-format C-DA on L2 performance. However, their assessment context covers comprehensive L2 performance, covering grammar, listening and pragmatics, etc. The effectiveness of C-DA on productive skills such as speaking and writing was not separately measured by employing the meta-analytic approach.

This meta-analysis is a fine-grained investigation on the effects of C-DA in promoting L2 learners' production. Specifically, it aims to synthesize existing empirical research, calculate the overall pooled effect size, conduct subgroup analyses and moderator analyses, eventually providing a comprehensive evaluation and offering practical guidance for future L2 teaching practices. The present study addresses the following research questions:

RQ1. What is the overall effect of C-DA on L2 production?

RQ2. How do the effects of C-DA vary across production types and measurement indicators?

RQ3. Do research duration, learners' L2 proficiency levels, and instructional contexts moderate the effects of C-DA on L2 production?

Methodology

Literature search

To ensure a comprehensive and precise literature search, a combination of various methods was employed, including systematic searches of authoritative databases, targeted manual searches of core journals identified through expert consultation and their recurrent citation in C-DA studies (Lu et al., 2026). English-language studies were retrieved from the Web of Science, the Education Resources Information Center (ERIC), APA PsycInfo, and ProQuest Dissertations and Theses. The Boolean search string was utilized: TS = ((“computerized dynamic assessment” OR “computer-mediated dynamic assessment” OR “mobile-mediated dynamic assessment”) AND (“second language production” OR performance OR “L2 writing” OR “EFL writing” OR “L2 speaking” OR “EFL speaking” OR complexity OR accuracy OR fluency)). Meanwhile, Chinese-language studies were retrieved from the China National Knowledge Infrastructure (CNKI). The search terms included “计算机动态评估” (computerized dynamic assessment), “动态评估” (dynamic assessment), “二语产出” (L2 production), “二语写作” (L2 writing), “英语写作” (EFL writing), “二语口语” (L2 speaking), and “英语口语” (EFL speaking). The reference lists of relevant published studies were manually searched to ensure the inclusion of pertinent and eligible studies that might have been missed in the earlier database searches. A total of 855 records from academic databases were identified.

Inclusion criteria

The inclusion criteria for the meta-analysis were as follows:

(1) Studies investigated the effects of C-DA on L2 production.

(2) Studies were original empirical research.

(3) Studies involved at least one task modality, either speaking or writing.

(4) Studies reported at least one dependent variable measuring EFL learners’ L2 production.

(5) Experimental design included a control group.

(6) Studies provided sufficient statistical data (e.g., sample size, mean, and standard deviation) to calculate effect sizes.

(7) Studies were published between 2014 and 2026. The year 2014 marked a turning point in the rapid development of educational technology (EdTech), providing the technical foundation for empirical studies on C-DA.

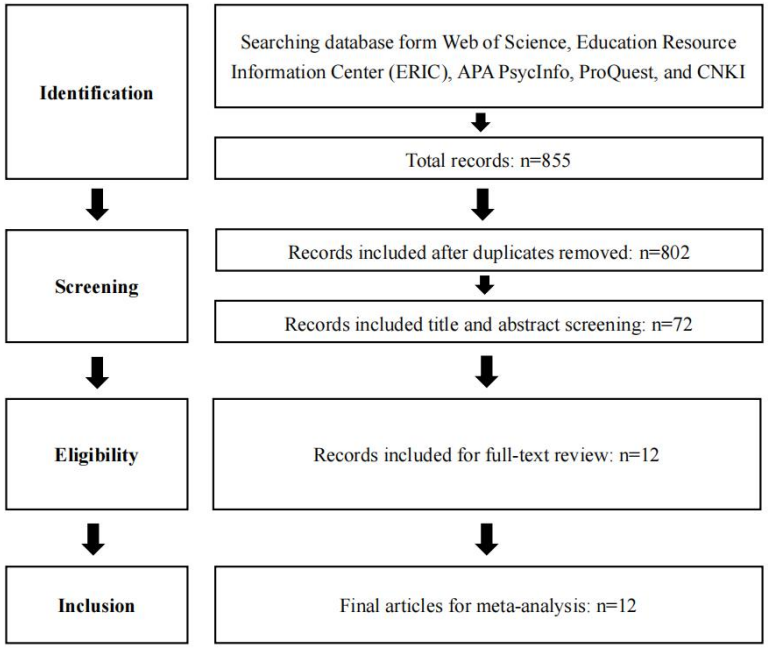


Figure 1. *PRISMA flow diagram*

Figure 1 presents a flowchart detailing the steps of literature search, screening and inclusion. The whole coding procedure followed the PRISMA guidelines to ensure transparency. Initially, a total of 855 records were identified through comprehensive and thorough searching. Subsequently, all retrieved records were imported into EndNote for duplicate removal and abstract screening, leaving 72 unique records for further screening. Then, a detailed full-text review based on inclusion criteria was conducted, leading to the exclusion of 21 records. After excluding original studies that failed to meet these criteria, 12 primary studies were ultimately selected as the final articles for this meta-analysis, involving 889 participants.

Coding

The coding scheme was developed based on our research purpose and the meta-analysis guidelines for L2 research proposed by Plonsky and Oswald (2015), which covered four main categories: (1) bibliographic information: author, year of publication and journal; (2) sample characteristics: sample size (for both experimental and control groups) and participant demographics (e.g., age, academic status, native language, and L2 proficiency); (3) methodological features: research design (e.g., C-DA implementation methods, intervention duration, and control group setup), task types (e.g., speaking or writing tests), and L2 production measures (e.g., complexity, accuracy, fluency, and overall scores); and (4) results: descriptive statistics and specific effect sizes.

Regarding the coding process, to ensure accuracy, two researchers coded the data independently. A high level of agreement was achieved between the independent

coders (Cohen's κ = 0.91). Any discrepancies during the coding process were resolved and a final decision was reached through sufficient discussion.

Data analysis

This study conducted a meta-analysis using Comprehensive Meta-Analysis 3.0 (CMA). First, effect sizes were extracted by entering the sample sizes, means, standard deviations, and other relevant data into CMA. Hedges's g was chosen as the effect size measure due to the small sample sizes present in this analysis (Lu et al., 2026). Following Cohen's (1988) benchmarks, values of 0.2, 0.5, and 0.8 were respectively viewed as the thresholds for small, medium, and large effects. The next step was to calculate the combined effect size and its confidence interval, so as to estimate the overall effect of C-DA on L2 production. Publication bias was subsequently assessed by conducting a comprehensive assessment using a funnel plot, the trim-and-fill procedure, Begg's rank correlation test, and Egger's regression test. Heterogeneity was also examined for the overall analysis and within subgroups. The results of heterogeneity tests determined the selection of appropriate meta-analytic models to compute pooled effect sizes. Finally, subgroup and moderator analyses were conducted to explore potential sources of variation across studies.

Results

Publication bias

This study combined the funnel plot with statistical tests to examine the risk of publication bias from multiple perspectives. Figure 2 shows that the distribution of effect sizes was not perfectly symmetrical, suggesting that there is a risk of publication bias. The funnel plot indicated a potential threat to the validity of meta-analytic findings. Three complementary statistical procedures were also conducted. Duval-Tweedie trim-and-fill iterative analysis demonstrated that the effect sizes before and after adjustment were identical, so that no missing studies needed to be imputed and no substantial publication bias was detected. Meanwhile, Begg's rank correlation test showed that Kendall's tau without continuity correction was 0.333 (p = 0.131). After continuity correction, the tau decreased to 0.318 (p = 0.150). Neither result reached statistical significance ($p > 0.05$), which showed a relatively balanced distribution of effect sizes and no significant association between effect size and standard error. Accordingly, Begg's test provided no clear evidence of publication bias. In contrast, the result of Egger's regression test (p = 0.009) indicated possible publication bias.

Egger's regression test yielded the evidence of possible publication bias, whereas the trim-and-fill procedure and Begg's rank correlation test did not support this finding. Rosenthal's fail-safe N, as a robustness check, was conducted and further showed that 358 additional studies with null results would be required to render the overall effect statistically nonsignificant. This number substantially exceeded the conventional threshold of 5k + 10 = 70 (k = 12), suggesting that publication bias is unlikely to have had a substantial influence on the meta-analysis. Taken together, these findings suggest that the overall conclusions were relatively reliable and stable.

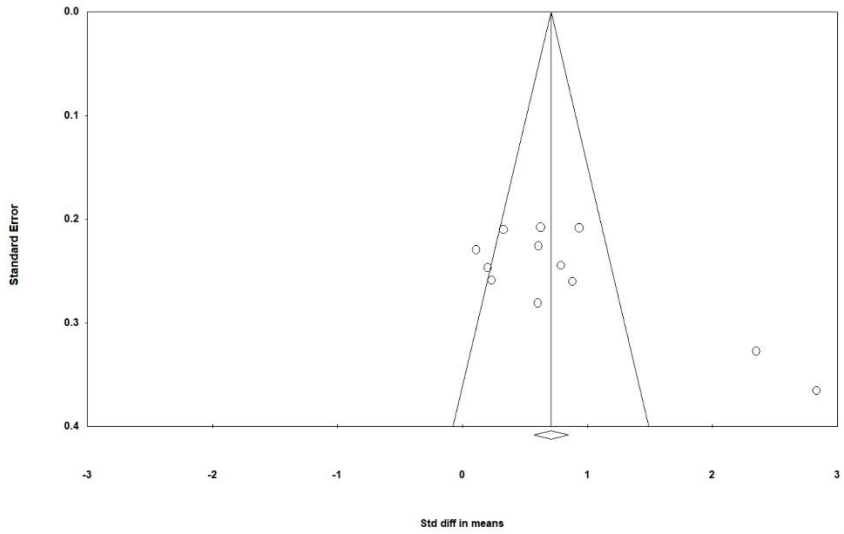


Figure 2. *Funnel plot of Hedges's g*

Overall Effect Size

The random-effects model was selected for the meta-analysis. The random-effects model allows the true effects to vary across studies and preserves more data in statistical estimation process, in contrast to fixed-effects models that assume homogeneity across the selected studies and samples (Borenstein et al., 2010). To evaluate the overall effect of C-DA on L2 production and examine its heterogeneity, a forest plot and effect-size analysis were generated. As shown in Figures 3 and 4, the pooled effect size was Hedges' g = 0.837 (95% CI [0.462, 1.212], $p < 0.001$), suggesting that computerized dynamic assessment exerts a substantial positive effect on L2 production. The heterogeneity test reflects varying degrees of heterogeneity among the independent samples (Shen et al., 2025). The analysis showed substantial heterogeneity (I^2 = 86.020%), reflecting that most of the variability across effect estimates was due to between-study heterogeneity rather than sampling error. In the context of this substantial heterogeneity, the study continued to adopt random-effects models to conduct subsequent subgroup analyses and moderator-effect tests for the sake of disclosing potential sources of heterogeneity.

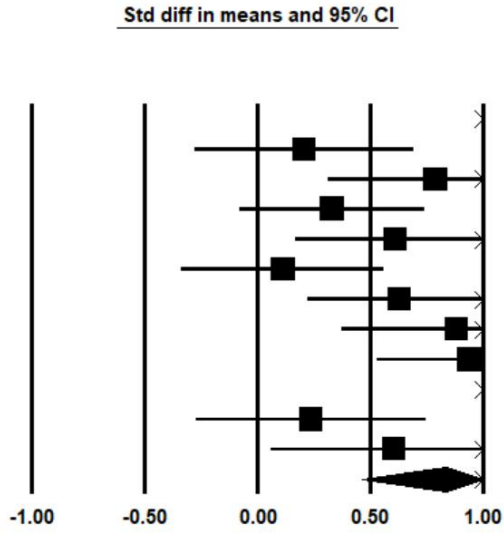


Figure 3. *Forest plot of overall effect size*

Model	Effect size and 95% confidence interval						Test of null (2-Tail)		Heterogeneity				Tau-squared			
Model	Number Studies	Point estimate	Standard error	Variance	Lower limit	Upper limit	Z-value	P-value	Q-value	df (Q)	P-value	I-squared	Tau Squared	Standard Error	Variance	Tau
Fixed	12	0.708	0.071	0.005	0.569	0.847	9.998	0.000	78.684	11	0.000	86.020	0.373	0.191	0.036	0.611
Random	12	0.837	0.191	0.037	0.462	1.212	4.377	0.000								

Figure 4. *Results of effect size analysis*

Subgroup Analysis

Effect sizes among different output types

This meta-analysis synthesized 12 empirical studies, one of which analyzed the effects of C-DA on both written and oral production. To compare the differential effects of C-DA on these two output modes, this study conducted a subgroup analysis. As illustrated in Table 1, C-DA had a positive effect on L2 writing production, with a pooled effect size of Hedges' g = 0.634 (95% CI [0.334, 0.935], $p < 0.001$), while the effect size for oral output was larger (g = 2.257, 95% CI [0.752, 3.762], $p = 0.003$). However, the I2 values of both subgroups exceeding 75% revealed the substantial within-group heterogeneity. In addition, the confidence interval for the speaking subgroup was considerably wider than that for the writing subgroup. This wider interval represented greater uncertainty surrounding the estimated effect on speaking, which may be attributed to the small sample sizes.

Table 1. *Subgroup analysis for different output types*

L2 production	k	Sample	*g*	*p*	95% CI		Heterogeneity	
					Lower	Upper	Q	I^2
Writing	10	764	0.634	< 0.001	0.334	0.935	37.085	75.732
Speaking	3	189	2.257	0.003	0.752	3.762	32.465	93.840

Effect sizes among different measurement indicators

Among the empirical studies included in this meta-analysis, three studies simultaneously employed multiple measurement indicators in their research design. Table 2 indicates that the effects of all four outcome measurement indicators were statistically significant ($p < 0.05$). Meanwhile, C-DA exerted differential effects on different measurement indicators. Its positive effects were more pronounced for fluency (g = 2.440, 95% CI [1.974, 2.905], $p < 0.001$) and accuracy (g =1.803, 95% CI [0.515, 3.091], $p = 0.006$), their effect sizes substantially exceeding those for overall score (g = 0.983) and complexity (g = 1.402). The overall score group (I^2 = 90.461%, Q = 94.346), the complexity group (I2 = 94.657%, Q = 56.145), and the accuracy group (I^2 = 92.288%, Q = 25.935) all showed high I2 values. This high level of heterogeneity may arise from pronounced differences in sample characteristics (e.g., L2 proficiency), or in measurement instruments. By contrast, heterogeneity in the fluency group was close to zero (I^2 = 0.000%, Q = 0.473), implying a high degree of consistency among the effect sizes. Nevertheless, this result warrants cautious interpretation, given the limited sample size in this subgroup.

Table 2. *Subgroup analysis for different measurement indicators*

Measurement indicators	k	Sample	*g*	*p*	95% CI		Heterogeneity	
					Lower	Upper	Q	I^2
Overall score	10	754	0.983	< 0.001	0.482	1.484	94.346	90.461
Complexity	4	298	1.402	0.015	0.272	2.532	56.145	94.657
Fluency	2	124	2.440	< 0.001	1.974	2.905	0.473	0.000
Accuracy	3	177	1.803	0.006	0.515	3.091	25.935	92.288

Moderator Analysis

Potential moderators were identified by reviewing previous relevant meta-analyses and considering the design and data characteristics of the included studies. Intervention duration, learners' L2 proficiency level, and instructional setting were selected as the three main moderators and moderator analyses were conducted using a random-effects model to evaluate the corresponding effect sizes. Table 3 presents the results.

Intervention duration was the first moderator to be examined. Two categories under Intervention duration were classified, including interventions lasting less than 15 weeks and interventions lasting more than 15 weeks. Studies with interventions lasting less than 15 weeks produced a statistically significant positive pooled effect of C-DA on L2 production (g = 1.014, 95% CI [0.537, 1.492], $p < 0.001$). For interventions lasting more than 15 weeks, the pooled effect was moderate and did not reach statistical significance (g = 0.502, 95% CI [−0.162, 1.165], $p = 0.138$). The results demonstrated that short-term interventions significantly promoted L2

production. However, the between-group difference was not statistically significant (Qbetween = 1.513, p = 0.219). The current evidence is insufficient to support intervention duration as a significant moderator of the effects of C-DA on L2 production.

Secondly, differences in effect-size estimates were observed across L2 proficiency levels. In the present study, participants' L2 proficiency was coded according to the *Common European Framework of Reference for Languages: Learning, Teaching, Assessment* (CEFR) (Council of Europe, 2020). As some Chinese studies employed locally defined proficiency indicators in the Chinese educational context, this meta-analysis uniformly mapped these Chinese academic indicators onto the CEFR framework for reporting participants' proficiency, ensuring data comparability. Two researchers independently completed the coding with a high level of inter-rater agreement. The statistical results illustrated that L2 learners at the B1 level (g = 0.911, 95% CI [0.343, 1.479], p = 0.002) and B2 level (g = 1.635, 95% CI [0.635, 2.636], p = 0.001), namely intermediate and upper-intermediate learners, had significant improvement in L2 production. The positive effects observed for these two proficiency levels were substantially larger than those for the A2 and C1 groups. Neither the A2 nor the C1 groups exhibited a statistically significant pooled effect. Nevertheless, the between-group difference in effect sizes across proficiency levels was not statistically significant (Qbetween = 4.328, p = 0.228). Hence, L2 proficiency can not be viewed as a significant moderator.

Finally, the last examined moderator in this meta-analysis was the instructional setting. Both the university group (g = 0.909) and the English-major group (g = 0.869) showed significant positive effects on L2 production. Effect sizes were largest for C-DA at the university level, whereas the intervention effect for the high school group was not statistically significant (g = 0.632, 95% CI [-0.371, 1.636], p = 0.217). Given that the test of between-group differences was not significant, instructional level was also not identified as a moderator of effect-size variation (Qbetween = 0.214, p = 0.898 > 0.05).

Table 3. *Results of moderator analysis*

Moderator	Type	k	*g*	*p*	95% CI		Q_b	$p(Q_b)$
					Lower	Upper		
Intervention duration	≤15weeks	8	1.014	< 0.001	0.537	1.492	1.513	0.219
	> 15weeks	4	0.502	0.138	-0.162	1.165		
L2 proficiency level	A2	1	0.329	0.635	-1.030	1.688	4.328	0.228
	B1	6	0.911	0.002	0.343	1.479		
	B2	2	1.635	0.001	0.635	2.636		
	C1	3	0.377	0.354	-0.420	1.174		
Instructional setting	High school	2	0.632	0.217	-0.371	1.636	0.214	0.898
	University	5	0.909	0.006	0.261	1.556		
	University (English major)	5	0.869	0.009	0.214	1.523		

Discussion

The Overall Effect of C-DA on L2 Production

This meta-analysis synthesized effect sizes from 12 experimental studies. The results evidently reveal that C-DA significantly influenced learners' written and oral L2 production. It is also broadly consistent with previous research, including Sherkuziyeva et al. (2023). This finding supports the view that C-DA enhances learners' task performance through strategic support when learners encounter difficulties, enabling them to apply strategies or knowledge that they could not deploy independently (Lu et al., 2026). Previous studies demonstrate that this approach may compensate for limitations in teacher-led assessment, such as limited time and delayed feedback. By providing structured assistance, C-DA, as a pedagogical strategy, can effectively activate the ZPD. Specifically, learners overcome the constraints of their current competence and gradually internalize the linguistic and cognitive resources required for more demanding activities (Wang, 2000). However, the substantial heterogeneity observed reflects that the estimated effects varied considerably. Such high heterogeneity indicates that even though C-DA effectiveness was generally robust across learners, the overall effects as a form of mediation in cognitive development are likely to be influenced by specific assessment requirements and instructional contexts.

Subgroup Effects of C-DA on L2 Production

The empirical studies coded in this meta-analysis employed diverse experimental designs. Researchers evaluated L2 production using multidimensional measures, including overall scores as well as complexity, accuracy, and fluency (CAF) indices, across different production types, namely writing and speaking. To address whether the effects of C-DA varied across these dimensions, two sets of subgroup analyses were conducted:

The first subgroup is written and oral production. Under the random-effects model, the speaking subgroup produced a larger pooled effect-size estimate than the writing subgroup. However, this difference should be interpreted with caution because the speaking subgroup contained relatively few studies. The automated corrective feedback provided by C-DA is immediate and targeted (Li et al., 2020), enabling L2 learners to draw on models and independently address pronunciation problems, grammatical errors, and pragmatic strategies within a relatively short period. These features may drive the immediate improvements in oral performance. Written production, by contrast, requires not only the correction and supplementation of linguistic knowledge, but also sustained cognitive processing to develop more elaborate content and rigorous logic. Xiao (2024) reported that computerized mediation was associated with reductions in technical, lexical, and discourse-level errors, whereas improvements in overall textual organization were less evident. Similarly, in Ren's (2019) intervention study, C-DA produced no significant improvements in lexical diversity, syntactic complexity, or fluency. Sherkuziyeva et al. (2023) disagree with the above viewpoints. They reported a significant posttest difference in writing complexity between the experimental and control groups, with the experimental group performing better. Overall, the available evidence suggests that C-DA does not affect all dimensions of written production equally. Such

variability may stem from the fact that writing tasks require the engagement of deeper cognitive structures (Xue & Pang, 2000). Meanwhile, the influence of C-DA on higher-order cognitive functions may require longer periods of mediation to make intervention effect become evident. Consequently, the effects of C-DA on writing have not yet demonstrated consistent patterns across studies, with the pooled effect remaining moderate.

The second subgroup concerns measurement indicators. C-DA produced positive effects across the examined measures of L2 production. The effects on fluency and accuracy were numerically larger than those on overall performance scores and complexity. As complexity, accuracy, and fluency interact dynamically and follow different developmental trajectories (Norris & Ortega, 2009), these three measures are not necessarily enhanced to the same extent. But accuracy and fluency can develop concurrently under some conditions (Yu, 2020). The data-analytic function of C-DA facilitates the statistical analysis of linguistic features in discourse, detecting spelling errors, grammatical and syntactic problems, and the frequency of cohesive devices. Its automated corrective feedback can therefore provide immediate and clear evaluations that lead learners to consolidate their linguistic knowledge and refine their expressions. Thus, C-DA can exert strong positive effects on fluency and accuracy. Overall scores and complexity are broader constructs influenced by multiple factors. Consequently, the cumulative effects on these measures require a longer period to emerge.

Effects of intervention duration, L2 proficiency level, and instructional setting

The between-group heterogeneity tests reveal that intervention duration, L2 proficiency level, and instructional setting did not significantly moderate the effects of C-DA on L2 production. Although no significant between-group differences were identified, the subgroup analyses still revealed meaningful intervention effects within individual subgroups.

First, studies with interventions lasting less than 15 weeks yielded a relatively large pooled effect size for C-DA, whereas the effects reported in studies lasting more than 15 weeks did not reach statistical significance. Effects of intervention duration on L2 production may be nonlinear. Cognitive load theory provides a plausible explanation for this finding. As a key component of instructional design, intervention duration, if inappropriately determined, may impose excessive extraneous cognitive load on learners (Paas et al., 2003). Well-designed short-term interventions help minimize such cognitive load while using the immediate feedback afforded by C-DA to increase learners' interest and self-efficacy.

Second, the intervention effects were statistically significant among learners at the B1-B2 levels (intermediate to upper-intermediate), with effect sizes exceeding those observed among learners at the C1 (advanced) and A2 (elementary) levels. This result is consistent with the meta-analytic findings of Shen et al. (2019). They reported that intermediate-level learners are more responsive to instructional interventions. Learners at the C1 level may already possess high L2 proficiency, so their room for further improvement becomes limited, making it difficult for C-DA interventions to overcome the "ceiling" of their language development. Learners at the elementary level represent the other extreme. On account of their inadequate linguistic knowledge, C-DA may primarily address gaps in basic language knowledge rather than facilitate deeper cognitive processing. Learners at the B1-B2

levels demonstrate greater developmental plasticity and are better positioned to internalize the benefits of the intervention.

Third, the university group showed the largest pooled effect size, followed by the university English-major group. The high school group also exhibited a positive effect, although it did not reach statistical significance. This ordering should be interpreted as a descriptive gradient based on effect sizes. The effectiveness of C-DA may depend on learners' autonomy and willingness to participate in exploratory learning (Lu et al., 2026). University students generally possess higher levels of metacognitive awareness and learner autonomy, which enables them to interpret and use the graduated prompts and immediate feedback provided by C-DA more effectively. Indeed, as Li and Mizumoto (2026) emphasize in the context of technology- and GenAI-mediated learning, fostering autonomous and efficacious tool use, rather than passive reliance, is critical for university EFL learners to maximize the developmental benefits of automated feedback (see also Teng, 2026). Furthermore, unlike EFL teaching in high school, university courses may have access to more favorable conditions for implementing C-DA, such as technological facilities, online platforms, and instructional organization. Sociocultural and pedagogical backgrounds need to be taken into account in C-DA implementation (Lu et al., 2026). These abundant pedagogical resources and advanced teaching facilities in the university allow their benefits to be fully leveraged. Previous studies conducted in university settings have primarily shown that C-DA can enhance learners' L2 performance (Shen et al., 2026; Yang & Qian, 2020) and may positively influence learning motivation, autonomous learning, and self-regulation (Estrada-Araoz et al., 2023).

Conclusion

This meta-analysis synthesized 12 empirical studies from Chinese and international literature to investigate the effects of C-DA on production and achieved three main findings. Firstly, C-DA had positive effects on both written and oral production, with a particularly large pooled effect for speaking. Secondly, the positive effects of C-DA were more pronounced for fluency and accuracy than other dimensions of measurement indicators. Thirdly, although intervention duration, learners' L2 proficiency, and educational level did not exert significant moderating effects on L2 production, all moderator analyses demonstrated that positive intervention effects of moderators varied in degrees. Specifically, relatively larger effects were observed for shorter interventions, intermediate and upper-intermediate learners, and university students. The results highlighted that appropriately designed short-term C-DA interventions may support the development of L2 production. Learners at intermediate and upper-intermediate levels also showed larger effect-size estimates than those at lower or advanced levels.

Our findings provide reference and implications for the wider use of C-DA in L2 educational environments. They may help teachers optimize instructional design, select assessment platforms suited to their pedagogical objectives, determine an appropriate intervention duration, and provide mediation aligned with EFL learners' ZPD. With the current trend toward the increasing integration of digital technology and L2 teaching, this study is also conducive to informing the future incorporation of AI-supported dynamic assessment tools.

Despite the comprehensive nature of this meta-analysis, several limitations should be noted. First and foremost, due to the limited number of studies involved, the educational backgrounds of learners were relatively restricted, which affects the generalizability of the findings to other cultural and educational contexts. In addition, some subgroups contained only a small number of studies or effect sizes. This reduced the statistical power to detect between-group differences and increased the uncertainty of the subgroup estimates. Therefore, the nonsignificant moderator results should be interpreted with caution. Finally, the analysis did not report the effects of C-DA on other L2 abilities and did not consider other potentially important moderators.

Several possible directions for future research can be drawn from the present study. Future meta-analyses should include a larger number of studies and learners from more diverse educational, linguistic, and cultural backgrounds, which contributes to increasing the statistical power of moderator analyses and ensuring a broader applicability of the results. Moreover, future research could also examine the effects of C-DA on reading, listening, translation, and other dimensions of language ability. Additionally, a wider range of potential moderators are recommended to be considered in future meta-analyses. Such work would provide a more comprehensive understanding of the conditions under which C-DA supports L2 development.

Funding:

This research received no external funding.

Informed Consent Statement:

Informed consent to participate was obtained from all individual participants included in the study.

Data Availability Statement:

The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest:

The authors declare that they have no competing interests.

Author's Contribution:

Both authors listed in the study have materially participated in the research and article preparation. Additionally, they have approved the final article. SH conceived of the presented idea and carried out the experiment. With support from MXL, SH performed the analysis, drafted the manuscript and designed the figures.

References

Babamoradi, P., Nasiri, M., & Mohammadi, E. (2018). Learners' attitudes toward using dynamic assessment in teaching and assessing IELTS Writing Task 1. *International Journal of Language Testing, 8*(1), 1–11.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic

introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods, 1*(2), 97–111. https://doi.org/10.1002/jrsm.12

Cai, J. (2012). The application of meta-analysis in second language research. *Foreign Language Teaching and Research, 44*(1), 105–115, 159–160.

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

Council of Europe. (2020). *Common European framework of reference for languages: Learning, teaching, assessment – Companion volume*. https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/16809ea0d4

Ebadi, S. & S. Vakili. (2019). Exploring EFL learners' developmental errors in academic writing through face-to-face and computer-mediated dynamic assessment. *Computer Assisted Language Learning, 35*(4). https://doi.org/10.1080/09588221.2019.1698616

Estrada-Araoz, E.G., Sayed, B.T., Niyazova, G.G., & Lami, D. (2023). Comparing the effects of computerized formative assessment vs. computerized dynamic assessment on developing EFL learners' reading motivation, reading self-concept, autonomy, and self-regulation. *Language Testing in Asia,* 13, 39. https://doi.org/10.1186/s40468-023-00253-1

Feng, S., & Zou, J. (2026). Group homogeneity enhances dynamic assessment efficacy in L2 English genitive learning: Preliminary evidence from integrating computerized and group DA. *Language Testing in Asia*, 16, Article 27. https://doi.org/10.1186/s40468-026-00447-3

Feuerstein, R., Rand, Y., & Rynders, J. E. (1988). *Don't accept me as I am: Helping "retarded" people to excel*. Plenum Press.

Geng, X., Meng, Y. R., & Du, W. B. (2022). Jisuanji dongtai pinggu de eryu fazhan shanbian he yanjiu kuangjia [Research on computerized L2 dynamic assessment: Development and framework]. Xiandai Waiyu [Modern Foreign Languages], 45(1), 126–136.

Ghahderijani, B. H., Namaziandost, E., Tavakoli, M., Kumar, T., & Magizov, R. (2021). The comparative effect of group dynamic assessment (GDA) and computerized dynamic assessment (C-DA) on Iranian upper-intermediate EFL learners' speaking complexity, accuracy, and fluency (CAF). *Language Testing in Asia*, 11, Article 25. https://doi.org/10.1186/s40468-021-00144-3

Guthke, J., & Beckmann, J. F. (2000). *The learning test concept and its applications in practice.* In C. S. Lidz & J. G. Elliott (Eds.), Dynamic assessment: Prevailing models and applications (pp. 17–69). Elsevier Science.

Imsa-ard, P., & Barrot, J. S. (2026). Combining multimodal technology-mediated and peer feedback: Effects on second language (L2) learners' complexity, accuracy, and fluency (CAF) in writing. *Computer Assisted Language Learning, 39*(4), 1023–1061. https://doi.org/10.1080/09588221.2024.2443770

Jiang, Z. (2026). Scaffolding early-stage simultaneous interpreting learning using computerized dynamic assessment. *Language Learning & Technology, 30*(1), 1–25. https://doi.org/10.64152/10125/73701

Lantolf, J. P., & Poehner, M. E. (2004). Dynamic assessment of L2 development: Bringing the past into the future. *Journal of Applied Linguistics, 1*(1), 49–72. https://doi.org/10.1558/japl.1.1.49.55872

Li, Y. (2015). A comparative study of feedback modes in English writing based on

dynamic assessment theory. *Foreign Language World*, (3), 59–67.

Li, Y., Li, Q., Huang, Y., et al. (2020). Embedding information technology in dynamic assessment in the college English speaking classroom. *The Chinese Journal of ICT in Education*, (12), 93–96.

Li, O., & Mizumoto, A. (2026). Measuring autonomous and efficacious GenAI use and GenAI reliance in EFL learning: Initial scale development and validation. *Applied Language Sciences*, *2*(1). https://doi.org/10.65553/ALS.260110

Lu, Q., Chen, M., Yang, L., Li, S., & Chen, Y. (2026). The effectiveness of computerized dynamic assessment in improving L2 performance: A three-level meta-analysis. *ReCALL*. Advance online publication. https://doi.org/10.1017/S0958344026100561

Lu, X., & Pan, L. (2026). Tailoring written feedback to fit: The role of cognitive learning styles in improving EFL writing with feedback. *Applied Language Sciences*, *2*(1). https://doi.org/10.65553/ALS.260112

Mardian, F., & Nafissi, Z. (2022). Synchronous computer-mediated corrective feedback and EFL learners' grammatical knowledge development: A sociocultural perspective. *Iranian Journal of Language Teaching Research, 10*(2), 115–136.

Meng, Y., Fu, H., & Wang, C. (2024). The effectiveness of computerized listening dynamic assessment: Attribute-based mediation model. *Language Learning & Technology, 28*(1), 1–28. https://doi.org/10.64152/10125/73580

Nicholas, A., Blake, J., Perkins, J., & Mozgovoy, M. (2026). Evaluating the effectiveness of a computerised dynamic assessment of L2 English email requests. *Computer Assisted Language Learning, 39*(1–2), 303–335. https://doi.org/10.1080/09588221.2024.2374775

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics, 30*(4), 555–578. https://doi.org/10.1093/applin/amp044

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38*(1), 1–4. https://doi.org/10.1207/S15326985EP3801_1

Pileh Roud, L. F., & Hidri, S. (2021). Toward a sociocultural approach to computerized dynamic assessment of the TOEFL iBT listening comprehension test. *Education and Information Technologies, 26*(4), 4943–4968. https://doi.org/10.1007/s10639-021-10498-z

Plonsky, L., & Oswald, F. L. (2015). *Meta-analyzing second language research.* In L. Plonsky (Ed.), Advancing quantitative methods in second language research (pp. 106–128). Routledge. https://doi.org/10.4324/9781315870908-6

Poehner, M. E., & Lantolf, J. P. (2013). Bringing the ZPD into the equation: Capturing L2 development during computerized dynamic assessment (C-DA). *Language Teaching Research, 17*(3), 323–342. https://doi.org/10.1177/1362168813482935

Qian, X., Meng, Y., & Yue, Z. (2023). Effects of dynamic-assessment-informed blended instruction on the development of academic English paraphrasing ability. *Foreign Language Education, 44*(2), 63–68. https://doi.org/10.16362/j.cnki.cn61-1023/h.2023.02.014

Qin, T., & van Compernolle, R. A. (2021). Computerized dynamic assessment of implicature comprehension in L2 Chinese. *Language Learning & Technology, 25*(2), 55–74. https://doi.org/10.64152/10125/73433

Ren, K. (2019). *Effects of using Juku Pigai in senior high school English writing instruction* [Master's thesis, Tianjin University].

Shen, J., Hou, Z., & Min, S. (2026). Exploring the use of chatbots in EFL argumentative writing: From the perspective of dynamic assessment. *International Journal of Applied Linguistics, 36*(1), 725–736. https://doi.org/10.1111/ijal.12809

Shen, Y., Zhang, L. J., & Zhou, L. (2019). A meta-analysis of empirical research on the effects of L2 listening strategy instruction. *Foreign Languages in China, 16*(6), 57–66. https://doi.org/10.13564/j.cnki.issn.1672-9382.2019.06.009

Shen, Y., & Zhang, L. J. (2020). Recent advances in meta-analysis in second language research in China and abroad, 2010 – 2019. *Foreign Language World*, (3), 70–79.

Shen, Y., Zhang, L. J., & Sun, Q. (2025). The effects of task complexity on L2 written production: A meta-analysis. *Modern Foreign Languages, 48*(1), 77–89. https://doi.org/10.20071/j.cnki.xdwy.20241030.006

Sherkuziyeva, N., Gabidullina, F. I., Ahmed Abdel-Al Ibrahim, K., & Bayat, S. (2023). The comparative effect of computerized dynamic assessment and rater-mediated assessment on EFL learners' oral proficiency, writing performance, and test anxiety. *Language Testing in Asia*, 13, Article 15. https://doi.org/10.1186/s40468-023-00227-3

Sternberg, R. J., & Grigorenko, E. L. (2002). *Dynamic testing: The nature and measurement of learning potential.* Cambridge University Press.

Tai, Z., & Wang, S. (2022). A review and prospects of dynamic assessment research abroad. *Journal of Tianjin Foreign Studies University, 29*(1), 84–95, 113.

Teng, M. F. (2026). Metacognition in using GenAI for language teaching and learning. *Emerging Educational Technologies*, *1*(1). https://doi.org/10.65553/EET.260101

Vakili, S., & Ebadi, S. (2022). Exploring EFL learners' developmental errors in academic writing through face-to-face and computer-mediated dynamic assessment. *Computer Assisted Language Learning, 35*(3), 345–380. https://doi.org/10.1080/09588221.2019.1698616

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Harvard University Press.

Wang, Q., Gao, Y., & Wang, X. (2024). The relationships between burnout, job satisfaction, and emotion regulation among foreign language teachers: A meta-analytic review. *Acta Psychologica*, 250, Article 104545. https://doi.org/10.1016/j.actpsy.2024.104545

Wang, W. (2000). Implications of Vygotsky's zone of proximal development theory for teaching reform in China. *Psychological Exploration,* (2), 17–20.

Xiao, Y. (2024). *Effects of dynamic assessment on the English writing performance of English majors in a blended-learning context* [Master's thesis, Jiangxi Normal University].

Xu, X. (2022). Construction and empirical evaluation of a big-data-driven dynamic assessment model for college English writing. *Journal of Jiangsu Vocational Institute of Commerce*, (5), 80–83.

Xue, Q., & Pang, W. (2000). Advances in research on the writing process from the perspective of modern cognitive psychology. *Psychological Science*, (4), 458–461,

465.
Yang, Y., & Qian, D. D. (2020). Promoting L2 English learners' reading proficiency through computerized dynamic assessment. *Computer Assisted Language Learning, 33*(5–6), 628–652. https://doi.org/10.1080/09588221.2019.1585882
Yu, H. (2020). Dynamic development of English oral fluency among Chinese learners: Interactions among complexity, accuracy, and fluency. *Foreign Language World*, (2), 81–89.

Author Bio

Han Shi is a doctoral student of Basic Education College, National University of Defense Technology, China. Her areas of interest are the development of EFL teachers, the integration of technology in EFL education, and the synergistic intersection of positive psychology with EFL education research. She focuses on frontier research in the aforementioned fields and actively participates in language teaching seminars. Her theoretical accomplishments and research capacity have been steadily improved.

Email: shihan@nudt.edu.cn

ORCID ID: https://orcid.org/0009-0003-0223-1769

Xiaolei Ma works as a Ph.D. supervisor in Basic Education College at National University of Defense Technology. He serves as an executive council member of the Language Governance Research Society of the Chinese Society for Language Modernization, a member of the Hunan Provincial Teaching Guidance Committee, a member of the Discipline Appraisal Group of the Hunan Provincial Academic Degrees Committee, and an executive council member of the Translators Association of Hunan. His research interests lie in technology-assisted language learning and foreign language education. His academic publications have appeared in a number of leading Chinese core journals (CSSCI indexed).

Email: maxiaolei@nudt.edu.cn

ORCID ID: https://orcid.org/0009-0003-6060-0980