

The effect of task repetition on L2 learners' writing performance: A meta-analysis

Language Teaching Research

1–27

© The Author(s) 2025

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/13621688251381268

journals.sagepub.com/home/ltr**Junge Liu** 

Beijing Normal University, China

Yinjie Tang 

Beijing Normal University, China

Abstract

This study systematically reviewed previous research on the effects of task repetition (TR) on second language (L2) learners' writing performance (WP) and conducted a meta-analysis to analyse the moderating variables. A total of 17 independent studies involving 442 participants were included. The results showed that TR could significantly enhance L2 learners' WP ($g = 0.25$, $p < .01$), though the effect size varied across different dimensions. Specifically, a significant positive medium effect size was found for fluency in L2 learners' WP ($g = 0.51$, $p < .01$), while accuracy and lexical complexity showed significant positive small effect sizes ($g = 0.24$ for both, $p < .01$ for both). In contrast, the positive effect on syntactic complexity was minimal and not statistically significant ($g = 0.06$, $p = .57$). Three variables including repetition interval, writing genre, and writing medium, exhibited distinct moderating effects. TR was found to be more effective with a one-week interval, in argumentative writing tasks, and when conducted in a computer-mediated environment. Although other factors, such as the repetition number, repetition type, language proficiency and age of the learners, and interaction style did not moderate the effect, they indicated general trends that offer valuable guidance for implementing TR. For instance, it can be argued that one repetition is sufficient to improve learners' WP and that TR may serve as a more efficient tool for adult learners. Overall, TR is an effective strategy for improving L2 learners' WP, regardless of their language proficiency, repetition type, or interaction style. The results of this study can serve as valuable guidance for L2 instructors in implementing TR effectively in their teaching practices. Given the current scarcity of research on TR's impact on L2 learners' WP, further studies are recommended to explore how TR can be optimized to enhance L2 learners' WP.

Keywords

accuracy, fluency, L2, lexical complexity, syntactic complexity, task repetition, writing performance

Corresponding author:

Junge Liu, School of Foreign Languages and Literature, Beijing Normal University, No.19, Xijiekouwai St, Haidian District, Beijing 100875, China

Email: jungeliu214@gmail.com

I Introduction

The impact of task repetition (TR) on bilingual performance is a prominent topic in Task-Based Language Teaching (TBLT) research (Luo & Xun, 2023). Bygate (2018) defines TR as the ‘repetition of a given configuration of purposes, and a set of content information’ (p. 2). Both task and repetition are argued to be essential for acquiring second language (L2) skills (DeKeyser, 2018). While numerous studies have investigated the role of TR in second language acquisition (SLA), particularly in oral production, and confirmed its benefits (e.g. Ahmadian, 2013; Hanzawa & Suzuki, 2023; Hunter, 2017; Nguyen et al., 2023; Róg & Krawiec, 2024; Sun & Liu, 2023; Thai & Boers, 2016), its impact on writing performance (WP) has only recently garnered significant attention (e.g. Abdi Tabari & Golparvar, 2024; Abdi Tabari, Khezlrou, & Ghanbar, 2024; Amiryousefi, 2016; J.C. Kim, 2024; Roothoof et al., 2022; Sánchez et al., 2020). However, findings on TR’s influence on WP are highly varied and sometimes contradictory.

In light of these diverse findings, this study seeks to evaluate the effect of TR on L2 learners’ WP within the CALF framework (i.e. syntactic complexity, accuracy, lexical complexity, and fluency) and identify key moderating factors through a meta-analysis.

II Literature review

I Theoretical background

This study drew on Kellogg’s (1996) writing model, Skehan’s (2014) Limited Attentional Capacity (LAC) hypothesis and skill acquisition theory (SAT) to provide a comprehensive framework for interpreting the findings of the meta-analysis. These frameworks collectively illuminate how TR may influence L2 learners’ WP by managing cognitive load, attentional resources, and the proceduralization of knowledge.

a Kellogg’s writing model and Skehan’s limited attentional capacity hypothesis. Kellogg’s (1996) writing model divides the writing process into three components: formulation, execution, and monitoring. The formulation process, which includes planning and translating, and the monitoring process, which involves reading and editing, place significant demands on the central executive: the system that manages the interaction between working memory and long-term memory. According to Ellis (2005), the execution process, which involves programming and performing writing tasks, can also be demanding for L2 learners due to challenges with typing or handwriting. As a result, all three processes impose substantial cognitive burdens on the central executive during writing.

Since the central executive has limited capacity, learners must prioritize one process over others. In TBLT, it is natural for learners to give priority to meaning, and thus the formal aspects of the languages are often overlooked. This aligns with Skehan’s (2014) LAC hypothesis, also known as the trade-off hypothesis in his earlier work, which highlights the difficulty of simultaneously balancing form and meaning in linguistic output due to limited attentional resources of L2 learners.

These theories support the role of TR in enhancing L2 learners' WP. Repeated tasks reduce the cognitive load on working memory by allowing learners to reuse prior content or procedural knowledge, freeing up cognitive resources to better balance meaning and form. Even without teacher instruction or feedback, repetition alone could possibly prompt learners to focus more on linguistic form while engaging in meaningful production (e.g. Jung, 2013).

b Skill acquisition theory. According to DeKeyser (2018), a key concept in SAT is that declarative knowledge can evolve into procedural knowledge. However, for procedural knowledge to be used efficiently and automatically, it must undergo a process of automatization, in which practice plays a crucial role. In the context of TR, the repetition of task content or procedure supports the transfer from declarative to procedural knowledge.

Proceduralization reduces the demands on working memory, as proceduralized information is stored as chunks, allowing for quicker retrieval. Therefore, through TR, cognitive load is reduced, freeing up cognitive resources and enabling L2 learners to access linguistic features more easily. As a result, learners are better able to achieve a balance across the CALF dimensions (Abdi Tabari, Khezrlou & Ghanbar, 2024).

2 The effect of TR on L2 learners' WP

Previous studies have investigated the effects of TR on L2 learners' WP using various indicators, such as holistic ratings (e.g. Hidalgo & Lázaro-Ibarrola, 2020), language-related episodes (LREs; e.g. Hidalgo & García Mayo, 2021; Hidalgo & Lázaro-Ibarrola, 2020; Y. Kim et al., 2020), cohesive features (e.g. Abdi Tabari et al., 2023). Among these, the CALF framework is the most commonly used and has been widely recognized as a reliable measure for capturing L2 learners' WP (e.g. Abdi Tabari & Golparvar, 2024; Abdi Tabari et al., 2023; Housen et al., 2012; López, 2018; Sánchez et al., 2020). However, findings on the effects of TR on CALF dimensions remain inconsistent.

Most studies suggested that TR could enhance at least one dimension of CALF in L2 learners' writing. For example, Abdi Tabari and Golparvar (2024) recruited 180 learners of English as a foreign language (EFL) with varying proficiency levels to complete two argumentative writing tasks at a one-week interval. Their study revealed consistent and positive improvements across all CALF measures, regardless of whether learners were informed in advance about the second task. Amiryousefi (2016) investigated 70 EFL learners' WP before and after TR and found that the TR group showed significant improvements in all fluency subdimensions and one subdimension of accuracy, although no notable improvements were observed in writing complexity. However, some research has claimed no significant correlation between TR and WP. For example, Sánchez et al. (2020) examined repeated performance on a decision-making task among 29 Spanish EFL learners in both oral and written modes. They found no significant effects on lexical complexity or accuracy, and observed a decline in syntactic complexity across both modalities. This decline was found to be more pronounced in writing. Moreover, fluency gains were only evident in the oral repetition group. The authors attributed these modality-dependent effects of TR to the temporal nature of writing. Unlike oral tasks, writing provides learners with 'a record of their language that they can look at and monitor'

(Nitta & Baba, 2014, p. 126). In other words, writers have more time to organize their thoughts and manage their attentional resources. As a result, competition among CALF dimensions may be less intense in writing, allowing writers to regulate these dimensions in a more organized and balanced way. This enables them to focus on both form and meaning during initial and repeated writing tasks, potentially reducing the impact of TR (Roothoof et al., 2022). The varying effects reflect the complex nature of TR's influence. Whether or not TR plays a role in WP and the magnitude of its effects appear to be dependent on a range of variables, which will be discussed in the next section.

3 *The moderators between TR and L2 learners' WP*

Previous studies have explored the moderators influencing the relationship between TR and L2 learners' WP from eight key perspectives: repetition interval, repetition number, repetition type, language proficiency of L2 learners, age of L2 learners, writing genre, interaction style (i.e. collaborative or individual writing), and writing medium (i.e. computer-mediated or pen-and-paper writing). Each of these factors will be discussed in turn.

a Repetition interval. In recent years, an increasing body of research has examined the effects of TR on L2 learners' language development from a cognitive psychology perspective, with a particular focus on the spacing between repetitions (Rogers, 2023). Advocates of immediate repetition emphasize the role of memory decay. For instance, Abdi Tabari et al. (2023) found that an interval of one week might have decreased the effects of repetition, suggesting that a shorter repetition cycle may better alleviate cognitive demands on learners. Similarly, Lambert et al. (2017) argued that only when tasks are repeated within a brief timeframe are the processes of formulation and conceptualization facilitated; otherwise, with memory decay, the benefits of TR may progressively diminish (Khezlrou, 2021). Some studies further suggest that consecutive repetitions over several days can positively impact complexity (e.g. Y. Kim et al., 2020), accuracy (e.g. Khezlrou, 2021; Y. Kim et al., 2020) and fluency (e.g. Khezlrou, 2021) in L2 learners' writing.

However, not all researchers agree on the effectiveness of immediate repetition for L2 learners. Hiver et al. (2024) observed that learners' WP followed a W-shaped progression, indicating that PR may take time to exert its full impact. Ellis et al. (2020) suggested that the interval between repetitions allows learners to confront the limitations of their initial performance, leaving traces that, though undeveloped initially, can be utilized in later attempts.

Although most previous studies have confirmed the importance of repetition intervals in the relationship between TR and L2 learners' WP, they often used weekly intervals without providing thorough justification (e.g. Abdi Tabari et al., 2023; Abdi Tabari, Khezlrou & Tian, 2024; Hidalgo & Lázaro-Ibarrola, 2020; Roothoof et al., 2022), typically due to research feasibility and alignment with teaching schedules. These varied intervals highlight the need for a meta-analytic approach to determine the optimal spacing for enhancing L2 learners' WP.

b Repetition number. TR helps L2 learners familiarize themselves with task content, procedures, or both within TBLT, thereby reducing working memory demands and allowing greater focus on language form. However, excessive TR is often associated with boredom and fatigue (Bygate, 2001). In a mini-meta-analysis, Bui and Yu (2021) concluded that even a single repetition can trigger internal readiness for the task, resulting in improved overall performance. Y. Kim et al. (2022) asked fifty-four Korean learners to complete two collaborative writing tasks and found that one repetition could improve students' writing fluency (total number of t-unit, $F = 39.99$, $p < .001$). Similarly, Gass et al. (1999) observed that morphosyntactic accuracy gains did not persist to Time 4,¹ which they attributed to learners' disinterest in the repeated tasks. Hidalgo and García (2021) also reported a decrease in LREs by Time 3² in the ER group, suggesting that one repetition was sufficient for learners to focus on language aspects.

Conversely, some researchers argued that TR did not positively impact learners' WP unless sustained over an extended period. For example, Abdi Tabari et al. (2023) found that performing tasks only twice did not affect the use of cohesive devices in L2 writing. J.K. Kim and Lee (2019) compared a two-repetition group, a three-repetition group, and a control group, finding that only the three-repetition group showed significant improvements over the control group in accuracy. Lambert et al. (2017) suggested that it may be necessary to perform a task up to five times so that students can effectively monitor their performance.

Up to now, little is known about the optimal number of repetitions required to maximize the benefits of TR for L2 learners' WP. Nitta and Baba (2014), in a 30-week study of 46 EFL (English as a foreign language) learners, found that competition among complexity, accuracy and fluency (CAF) was dynamic. A single repetition might initially draw learners' attention to fluency, but over time, the focus shifted to lexical and syntactic dimensions as learners challenged themselves to incorporate more varied structures and vocabulary.

c Repetition type. According to Patanasorn (2010), TR can be classified into three types: exact repetition (ER), which involves the same content and procedure; procedural repetition (PR), which involves the same procedure but different content; and content repetition (CR), which involves the same content but a different procedure.

Previous studies have shown that different types of repetition can have varied effects on different dimensions of L2 learners' WP. For example, Y. Kim et al. (2020) found that ER positively influenced syntactic complexity (morphemes/t-unit, $t = 2.14$, $p = .04$) and target-feature accuracy ($t = 3.33$, $p = .003$) in the writing of Korean learners, although some subdimensions showed no significant differences. In contrast, the PR group demonstrated a notable decrease in both syntactic complexity (morphemes/t-unit, $t = 7.03$, $p < .001$) and target-feature accuracy ($t = 2.65$, $p = .019$) in the repeated task. Amiryousefi's (2016) study found that both ER and PR could enhance EFL learners' written production, although the ER group outperformed the PR group in fluency (e.g. number of words, $t = 13.344$, $p = .000$) and one dimension of accuracy (percentage of error-free clauses, $t = 2.251$, $p = .028$). The varying effects of PR may be attributed to learners' differing levels of familiarity with tasks 1 and 2. Itze Arredondo-Tapia and Emmanuell Garcia-Ponce (2021) suggested that while cognitive resources were freed as learners focused less on

procedural aspects, this may not fully offset the disadvantage introduced by unfamiliarity with task content, leading to unpredictable fluctuations in fluency and accuracy.

Because CR involves a different task procedure, it is often used to investigate the effects of various interventions on L2 learners' WP, such as task sequencing (e.g. Abdi Tabari, Khezrlou & Ghanbar, 2024), types of feedback (e.g. Roothoof et al., 2022) and the timing and availability of instruction (e.g. Khezrlou, 2021). We will not involve this type of repetition in our meta-analysis.

d Language proficiency of L2 learners. Previous studies have suggested that L2 learners' language proficiency can be a significant predictor of the effect of TR on their WP. Higher proficiency L2 learners are believed to possess a greater ability to encode linguistic elements, thus they might have more attentional resources, which enhance their ability to conceptualize task content (Lambert et al., 2017). This ability allows high-proficiency learners to focus more on the formal aspects of language, resulting in a more balanced development across the CALF dimensions. In contrast, low-proficiency learners may struggle to allocate attention across all aspects of language due to limited attentional resources and language knowledge, even in repeated tasks.

Nitta and Baba (2014) noted that low-proficiency learners tended to prioritize fluency at the expense of lexical and syntactic elements. Y. Kim et al. (2020) found that low-proficiency L2 learners had difficulty using multiple clauses per t-unit, showing no significant improvement in syntactic complexity during repeated tasks (clauses/t-unit, $t = 1.72$, $p = .10$). Similarly, Y. Kim et al. (2022) analysed the collaborative writing of beginner-level Korean learners and found that TR benefited their writing fluency (total number of t-unit, $F = 39.99$, $p < .001$) but not their syntactic complexity, due to their limited knowledge of complex syntax (clauses/t-unit, $F = .28$, $p = .87$). Amiryousefi (2016) highlighted the benefits of TR for low-proficiency learners, explaining that as learners become more familiar with the task, additional attentional and processing resources become available, enabling them to optimize their performance in subsequent tasks. However, no significant improvement in syntactic complexity was found in either the ER (dependent clauses/total clauses, $t = 1.48$, $p = .148$) or PR (dependent clauses/total clauses, $t = .54$, $p = .410$) group. The trade-off between fluency and accuracy (AF) on one hand, and complexity on the other, can be attributed to low-proficiency learners prioritizing AF during writing tasks. Complexity involves higher cognitive demands, which are typically addressed once AF has become automatized and consolidated. Due to their limited attentional resources, these learners may overlook the complexity of their writing. From the perspective of SAT, more complex skills can only be effectively tackled once simpler skills have become automatic.

The relationship between L2 learners' language proficiency and their WP, however, is not always straightforward. Some studies have suggested a more complex relationship. Huh et al. (2018), for example, examined the writing development of four students over time from a dynamic systems theory perspective and found that L2 learners with different proficiency levels exhibited distinct developmental trajectories. Liu (2024) confirmed this dynamic relationship, showing that lower-proficiency learners initially performed better in inter-sentential cohesion, while intermediate-level learners excelled later in the repetition cycle. Sánchez et al. (2020) argued that while TR effects are not

directly moderated by proficiency, there is a complex relationship, with some dimensions. For example, lexical complexity is proficiency-dependent while others are not. López (2018) found that L2 learners at different proficiency levels performed similarly across writing tasks, with proficiency's moderating effect emerging only when feedback was provided in subsequent tasks.

e Age of L2 learners. Most previous studies on the impact of TR on WP have focused primarily on adult learners (e.g. Abdi Tabari, Khezlrou & Ghanbar, 2024; Amiryousefi, 2016; Khezlrou, 2021; Y. Kim et al., 2020, 2022; Qiu & Lo, 2017; Sánchez et al., 2020), while research focusing on young learners (YLS) remains relatively scarce (e.g. Hidalgo & García Mayo, 2021; López, 2018, p. 57; Roothoof et al., 2022; Sánchez et al., 2020), and has produced varied results. For instance, Roothoof et al. (2022) examined WP of 75 L2 learners aged 10–12 years and found that TR led to only minimal improvements in one measure of complexity (mean length of clause, $p < 0.05$) within the CAF framework and there was a significant decrease for the error-free clause ratio ($p < 0.05$). In contrast, Hidalgo and García Mayo (2021) argued that TR is a valuable tool for enhancing YLS' (aged 11–12 years) attention to language form during collaborative writing tasks. According to Ullman's (2006) declarative and procedural (DP) model, children predominantly rely on procedural memory, while adults depend more on declarative memory. Declarative knowledge, which is easier to be transferred to subsequent practice, is supported by declarative memory. Therefore, it can be argued that, compared to YLS, adult learners are better able to quickly acquire declarative knowledge from prior practice, allowing them to allocate more attentional resources to language form in subsequent writing tasks. However, in fact, few studies have compared the impact of TR on WP across different age groups, and most of those associated them with different levels of language proficiency.

f Writing genre. Writing genre has been shown to explain a significant amount of variance in L2 learners' WP. Among the various genres, the distinction between descriptive and argumentative essays is one of the most widely studied topics. Many scholars have confirmed the differing effects of these genres on L2 learners' WP. Fukunaga (2023) investigated the impact of descriptive and argumentative essays on Japanese learners' writing across CAF dimensions over a period of 16 weeks. The study found that both genres improved in complexity and accuracy, but the effect size for descriptive essays was larger than that for argumentative essays. However, no significant improvements in accuracy were observed in the descriptive essays. In contrast, the study by Abdi Tabari, Khezlrou and Ghanbar (2024) indicated that TR could significantly improve accuracy of both genres ($p = .00, .00$). However, significant fluency improvements were observed only in the argumentative genre ($t = 4.55, p = .00$).

g Interaction style. The writing tasks in TBLT can be performed either individually or collaboratively. Roothoof et al. (2022) found that TR plays a minor role in improving students' writing output, with effects observed only at the complexity level ($p < .05$). This finding diverges from results reported in other studies (e.g. Hidalgo & Lázaro-Ibarrola, 2020; Y. Kim et al., 2020). They attributed this discrepancy to the differences between pair work and individual work. In collaborative writing, learners are encouraged to engage

in discussion and negotiation and assist each other in noticing both the meaning and form of their writing. LREs recorded during these discussions have successfully demonstrated this effect (e.g. Hidalgo & Lázaro-Ibarrola, 2020; Y. Kim et al., 2020, 2022). However, few studies have considered the interaction style as an important moderator or compared its impact to that of individual writing in TR.

h Writing medium. Most previous studies have used pen-and-paper as the medium for writing tasks. However, with the growing integration of technology in education, the use of computers in language learning will become more and more prevalent. Amiryousefi (2016) examined 70 Iranian EFL learners and found that TR had beneficial effects on their task-based, computer-mediated L2 written production. Similarly, Abdi Tabari and Golparvar (2024) asked 180 EFL learners to complete two writing tasks using a word processor and found a consistent and positive enhancement across all linguistic measures in the repeated task. Despite these findings, the effects of TR on L2 learners' WP through computer-mediated versus pen-and-paper mediums remain unclear and warrant further investigation.

In summary, the effect of TR on L2 learners' WP remains divergent and has yet to reach a consensus. Through a meta-analysis of empirical studies, this research aims to address the following research questions:

- Research question 1: What is the effect of task repetition on L2 learners' WP within the framework of CALF?
- Research question 2: What are the moderating variables that influence the effect of task repetition on L2 learners' WP?

III Method

I Initial search and screen

To identify relevant studies for this meta-analysis, a comprehensive literature search was conducted on 24 October 2024. The search string was developed based on a review of the literature concerning the impact of TR on L2 learners' WP. The search string was defined as: ('task repetition' OR 'procedural repetition' OR 'content repetition' OR 'exact repetition' OR 'task iteration') AND ('L2 writing' OR 'second language writing' OR 'L2 written' OR 'collaborative writing' OR 'writing task') AND ('CAF' OR 'complexity' OR 'fluency' OR 'accuracy' OR 'lexical complexity' OR 'CALF' OR 'syntactic complexity' OR 'syntactic sophistication' OR 'lexical sophistication') AND ('second language acquisition' OR 'SLA' OR 'second language learning' OR 'L2' OR 'EFL' OR 'ESL' OR 'second language'). Seven databases were searched for the targeted studies: ProQuest, Elsevier, Wiley Online Library, EBSCO, Sage, Web of Science, and Springer Link. Only English-language, peer-reviewed, and accessible studies were included. After the preliminary screening, a total of 453 records were retrieved from the databases.

2 Selection criteria

The selection criteria for this meta-analysis were established to ensure the inclusion of high-quality, relevant studies specifically investigating the effects of TR on L2 learners' WP. The specific inclusion criteria are as follows:

1. The research topic must focus on the impact of TR on L2 learners' WP. Studies that incorporate external interventions, such as teacher feedback and instructional guidance, will be excluded.
2. The included studies must include at least one dimension of the CALF framework. Studies that solely examine the impact of TR on cohesive features in L2 writing, explicit or implicit knowledge acquisition, or other evaluative measures, such as holistic ratings, LREs, or verb argument construction-based indices, will be excluded.
3. The included studies must report the necessary data (e.g. sample size, mean, standard deviation, etc.) to calculate the effect sizes. Studies that describe dynamic changes without providing specific data will be excluded.
4. The writing process in the included studies must involve a complete process of self-construction of meaning. Studies that focus only on sentence-level writing or dictogloss tasks will be excluded.
5. The included studies must be empirical, employing repeated measurements or comparisons between groups.
6. To ensure the quality and reliability of this meta-analysis, only studies published in SCI or SSCI journals, book chapters, or doctoral theses will be included.

The inclusion criteria were deliberately stringent to ensure the quality of the meta-analysis. Studies that did not meet any of the criteria were excluded. After a thorough screening based on these criteria, a total of 10 studies, comprising 442 participants, met all inclusion criteria and were incorporated into the meta-analysis. Among these, two are doctoral theses, one is a chapter from a book, and the remaining studies were published in SCI or SSCI journals. The publication years of these studies range from 2016 to 2024.

3 Coding

To ensure coding accuracy, the researcher conducted two coding sessions in October and November 2024. The coding for this study consisted of four main parts. First, publication features were recorded, including the title, author(s), publication year, and journal or book information. Second, substantive features were documented, covering basic information about the participants (e.g. age, second language proficiency), as well as details about the study design (e.g. sample size, repetition type, repetition interval, repetition number, writing genres, writing medium, and interaction style). Third, methodological features were recorded, specifically the dimensions used to measure L2 learners' WP. Fourth, outcome features were documented, which included the mean, standard deviation (SD), and t-values for the CALF measures. If a single paper reported more than one independent study, each study was coded separately. In total, 17 independent studies were recorded. The detailed information of each included study is presented in Table 1.

Table 1. Coding of the included studies.

Studies	N	Repetition interval	Repetition number	Repetition type	Language proficiency	Age	Writing genre	Interaction style	Writing medium	Evaluation dimensions
Amiryousefi (2016) (1)	35	1w	4	ER	low	Adult	narrative	individual	computer	CAF
Amiryousefi (2016) (2)	35	1w	4	PR	low	Adult	narrative	individual	computer	CAF
Mehrang (2016)	24	4w	1	ER	intermediate	14–50	problem solving	individual	pen	CALF
López (2018) (1)	7	8d	1	ER	intermediate	Adult	decision making	individual	pen	CALF
López (2018) (2)	8	8d	1	ER	low	YL	decision making	individual	pen	CALF
Hidalgo and Lázaro-Ibarrola (2020)	20	1w	2	ER	low	YL	narrative	collaborative	pen	CALF
Y. Kim et al. (2020) (1)	23	1d	1	ER	low	Adult	narrative	collaborative	pen	CA
Y. Kim et al. (2020) (2)	15	1d	1	PR	low	Adult	narrative	collaborative	pen	CA
Sánchez et al. (2020) (1)	8	8d	1	ER	low	YL	decision making	individual	pen	CALF
Sánchez et al. (2020) (2)	7	8d	1	ER	intermediate	Adult	decision making	individual	pen	CAF
Y. Kim et al. (2022)	29	1d	1	ER	low	Adult	narrative	collaborative	pen	CAF
Roothoof et al. (2022)	21	1w	2	ER	low	YL	narrative	individual	pen	CALF
Abdi Tabari and Golparvar (2024) (1)	30	1w	1	ER	low	Adult	narrative	individual	computer	CALF
Abdi Tabari and Golparvar (2024) (2)	30	1w	1	ER	intermediate	Adult	narrative	individual	computer	CALF
Abdi Tabari and Golparvar (2024) (3)	30	1w	1	ER	high	Adult	narrative	individual	computer	CALF
Abdi Tabari, Khezrlou and Ghanbar (2024) (1)	60	1w	1	ER	high	Adult	narrative	individual	pen	CA
Abdi Tabari, Khezrlou and Ghanbar (2024) (2)	60	1w	1	ER	high	Adult	argumentative	individual	pen	CA

Notes. ER = exact repetition; PR = procedural repetition; YL = young learner; CALF = syntactic complexity, accuracy, lexical complexity, and fluency; CAF = complexity, accuracy, and fluency.

4 Data extraction

Hedges's g was used as the effect size indicator due to the small sample sizes of the studies included ($n < 60$) (Durlak, 2009). Almost all of the included studies employed multiple measures to evaluate each dimension of CALF. However, the unit of meta-analysis is the individual research study (Lipsey & Wilson, 2001). Including more than one effect size from a single study would 'violate the assumption of independent data points' and 'render the statistical results highly suspect' (p. 105). Therefore, in order to avoid the 'sample size inflation' effect (Li, 2010, p. 327) and to adhere to the core principle of meta-analysis, i.e. 'one study, one effect size', this study first calculated effect size for each measure and then combined them into a composite effect size for meta-analysis. Specifically, we first used the mean, standard deviation, and sample size extracted from the included studies to calculate the effect size Hedges's g using an online calculator (<https://www.campbellcollaboration.org/calculator/d-means-sds>). For studies that did not report the statistical information mentioned above, we used t -values to calculate Hedges's g (e.g. Hidalgo & Lázaro-Ibarrola, 2020). In total, we got 138 Hedges's g . Then, we combined different measures of one independent variable into one effect size. After that, the effect sizes, confidence intervals, and sample sizes were entered into the Comprehensive Meta-Analysis 3.7.

To exclude outliers, we conducted a sensitivity analysis using the 'one study removed' method. According to Li (2010), any study that resulted in a mean effect size change greater than 0.05 should be excluded. Since 'one outlier in one analysis may not necessarily be an outlier in another analysis' (p. 329), we conducted five separate analyses and ultimately removed three studies (i.e. Y. Kim et al. (2020) (2) was removed from total effect, Roothoof et al. (2022) was removed from syntactic complexity and accuracy).

5 Publication bias test

The funnel plot showed that most of the effect sizes were concentrated in the middle and upper part of the funnel plot. Both sides of the plot were approximately symmetrically uniformly distributed, indicating that there was no significant publication bias in the present study (see Figure 1). Further results from the publication bias test showed that the coefficient of insecurity was 339, which was well above the critical value of 90 ($k = 16$), suggesting that the robustness of this meta-analysis was strong. Additionally, Egger's regression analysis revealed no evidence of publication bias in the studies ($p = .51 > .05$).

6 Heterogeneity test

The results of the heterogeneity test (see Table 2) indicated significant heterogeneity in the total effect ($I^2 = 65.76, p < .01$). Among the different dimension of CALF, syntactic complexity ($I^2 = 84.10, p < .01$), accuracy ($I^2 = 44.13, p < .05$) and fluency ($I^2 = 51.58, p < .05$) exhibited significant heterogeneity, necessitating the use of a random-effects model (Borenstein, 2009). However, for lexical complexity, there was no

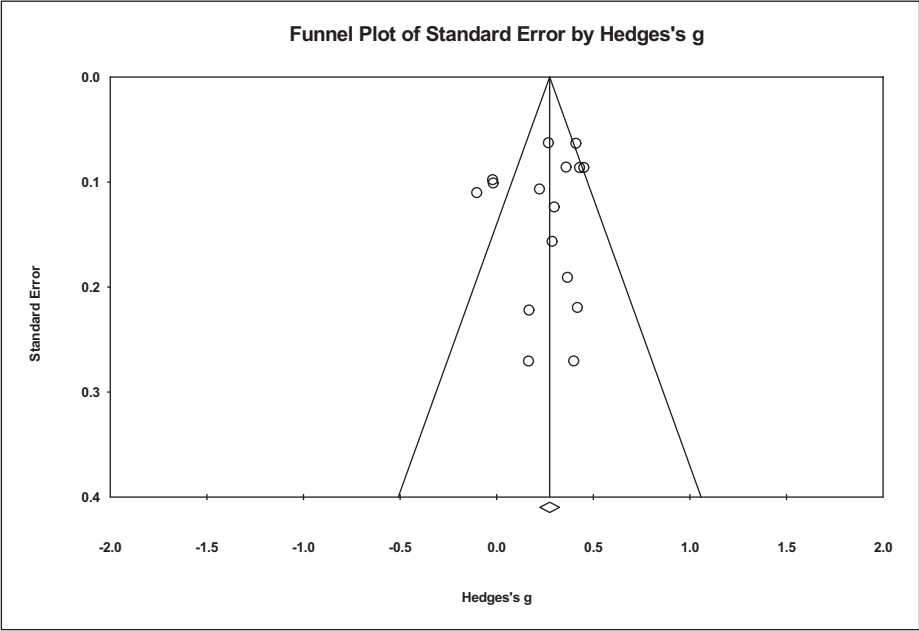


Figure 1. Funnel plot of the included studies.

Table 2. Heterogeneity test.

Studies		Heterogeneity				Tau squared	
		Q	df(Q)	p	I ²	Tau squared	Tau squared
Syntactic complexity	16	94.32	15	0.00	84.10	0.12	0.06
Accuracy	16	26.85	15	0.03	44.13	0.04	0.03
Lexical complexity	12	15.29	11	0.17	28.05	0.02	0.02
Fluency	15	28.91	14	0.01	51.58	0.07	0.05
Total	16	43.81	15	0.00	65.76	0.02	0.01

substantial heterogeneity ($I^2 = 28.05$, $p > .05$), so we chose a fixed model to calculate the effect size.

IV Results

I The effect of TR on L2 learners' WP

This study selected Hedges's g as the effect size indicator. Following the effect size measurement standards proposed by Cohen (1988), values of 0.2, 0.5, and 0.8 were used

Table 3. The effect of task repetition (TR) on second language (L2) learners’ writing performance (WP) across syntactic complexity, accuracy, lexical complexity, and fluency (CALF).

	Studies	N	<i>g</i>	SE	95% CI		<i>Z</i>	<i>p</i>
					LL	UL		
Syntactic complexity	16	421	0.06	0.10	−0.14	0.25	0.57	0.57
Accuracy	16	421	0.24	0.08	0.09	0.38	3.14	0.00
Lexical complexity	12	305	0.24	0.06	0.13	0.35	4.28	0.00
Fluency	15	404	0.51	0.10	0.31	0.70	5.16	0.00
Total	16	427	0.25	0.05	0.16	0.35	5.21	0.00

Notes. CI = confidence interval; LL = lower limit; UL = upper limit.

as the critical thresholds for small, medium, and large effects. Table 3 shows that TR has a positive effect on L2 learners’ WP and the effect is statistically significant ($g = 0.25, p < .01$). However, the impact varies across different dimensions of CALF. Specifically, TR has a medium effect on fluency ($g = 0.51, p < .01$) and small effects on lexical complexity ($g = 0.24$) and accuracy ($g = 0.24$), with all effects being statistically significant ($p < .01$). However, the positive effect of TR on syntactic complexity was not significant, indicating a very small effect ($g = 0.06, p = .57$).

2 Moderating variables between TR and L2 learners’ WP

This study conducted a subgroup analysis to assess the moderating effect of various variables on the relationship between TR and L2 learners’ WP.

Due to the scarcity of studies examining the impact of TR on L2 learners’ WP, only four periods of intervals were included in the analysis. Nevertheless, the limited studies provided insights into the trends of TR’s effects on L2 learners’ WP. Table 4 showed that repetition interval could serve as a significant moderator in the relationship between TR and WP, with notable differences observed across all dimensions of CALF and the total effect. It can be argued that a one-week interval is the most beneficial for L2 learners’ WP, exhibiting positive effects across all dimensions of CALF. Conversely, when the repetition interval was extended to four weeks, the effect of TR on L2 learners’ WP became negative.

Table 5 indicates that the number of repetitions did not moderate the relationship between TR and WP, as no dimension showed a significant difference. However, the table suggested that when the task was repeated once, all dimensions of CALF improved. When the number of repetitions was increased to four, the positive effect of TR on L2 learners’ WP, with the exception of fluency (while $p > .05$), showed no significant further improvement. It can be argued that one repetition is sufficient to enhance L2 learners’ WP.

Table 6 shows that most previous studies have primarily focused on ER writing tasks, while there is a notable lack of exploration into PR writing tasks. The moderating effect

Table 4. Moderating effect of repetition interval.

	Sub-moderators	N	g	95% CI		Z	p	Q
				LL	UL			
Syntactic complexity	1d	67	-0.45	-1.75	0.86	-0.67	0.50	8.04*
	1w	300	0.27	0.13	0.41	3.73	0.00	
	8d	30	-0.04	-0.21	0.14	-0.41	0.68	
	4w	24	0.00	-0.57	0.57	0.00	1.00	
Accuracy	1d	67	0.01	0.16	0.03	-0.32	0.33	8.03*
	1w	300	0.34	0.07	0.01	0.20	0.48	
	8d	30	0.25	0.28	0.08	-0.30	0.81	
	4w	24	-0.15	0.19	0.04	-0.52	0.23	
Lexical complexity	1w	251	0.35	0.07	0.00	0.22	0.48	10.34**
	8d	30	-0.06	0.12	0.01	-0.28	0.17	
	4w	24	-0.04	0.28	0.08	-0.58	0.51	
Fluency	1d	29	0.69	0.26	1.20	2.65	0.01	15.05**
	1w	321	0.46	0.11	0.68	4.11	0.00	
	8d	30	-0.14	0.20	0.25	-0.70	0.48	
	4w	24	0.79	0.15	1.08	5.35	0.00	
Total	1d	52	0.29	0.10	0.49	3.03	0.00	19.52**
	1w	321	0.35	0.29	0.41	11.42	0.00	
	8d	30	0.12	-0.08	0.32	1.14	0.25	
	4w	24	-0.10	-0.32	0.11	-0.93	0.35	

Notes. CI = confidence interval; LL = lower limit; UL = upper limit. * $p < .05$; ** $p < .01$.

Table 5. Moderating effect of repetition number.

	Sub-moderators	N	g	95% CI		Z	p	Q
				LL	UL			
Syntactic complexity	1	331	0.09	-0.13	0.31	0.78	0.43	5.60
	2	20	-0.63	-1.20	-0.05	-2.15	0.03	
	4	70	0.09	-0.13	0.31	0.80	0.43	
Accuracy	1	331	0.21	0.09	0.01	0.02	0.39	0.96
	2	20	0.40	0.29	0.08	-0.16	0.96	
	4	70	0.33	0.12	0.01	0.10	0.56	
Lexical complexity	1	264	0.23	0.06	0.00	0.12	0.34	0.45
	2	41	0.45	0.32	0.11	-0.18	1.09	
Fluency	1	293	0.52	0.11	0.73	4.65	0.00	1.63
	2	41	0.28	0.25	0.77	1.12	0.26	
	4	70	1.14	0.70	2.51	1.64	0.10	
Total	1	316	0.26	0.15	0.37	4.50	0.00	0.25
	2	41	0.17	-0.17	0.50	0.97	0.33	
	4	70	0.25	0.05	0.44	2.48	0.01	

Notes. CI = confidence interval; LL = lower limit; UL = upper limit.

Table 6. Moderating effect of repetition type.

	Sub-moderators	N	g	95% CI		Z	p	Q
				LL	UL			
Syntactic complexity	Exact repetition	371	0.20	0.07	0.32	3.14	0.00	1.07
	Procedural repetition	50	-0.96	-3.13	1.22	-0.86	0.39	
Accuracy	Exact repetition	371	0.25	0.09	0.01	0.08	0.42	0.18
	Procedural repetition	50	0.18	0.15	0.02	-0.11	0.46	
Lexical complexity	Exact repetition	305	0.24	0.06	0.00	0.13	0.35	0.00
	Exact repetition	369	0.50	0.10	0.70	4.86	0.00	0.14
Fluency	Procedural repetition	35	0.64	0.35	1.32	1.81	0.07	
	Exact repetition	392	0.26	0.15	0.36	4.90	0.00	0.08
Total	Procedural repetition	35	0.22	0.01	0.43	2.08	0.04	

Notes. CI = confidence interval; LL = lower limit; UL = upper limit.

Table 7. Moderating effect of the language proficiency of second language (L2) learners.

Sub-moderators		N	G	95% CI		Z	p	Q
				LL	UL			
Syntactic complexity	high	150	0.38	0.27	0.49	6.66	0.00	10.12**
	intermediate	68	0.11	−0.11	0.33	0.95	0.34	
	low	203	−0.12	−0.48	0.24	−0.68	0.50	
Accuracy	high	150	0.31	0.11	0.01	0.09	0.54	0.36
	intermediate	68	0.19	0.23	0.05	−0.27	0.65	
	low	203	0.24	0.10	0.01	0.04	0.43	
Lexical complexity	high	150	0.29	0.08	0.01	0.14	0.45	0.99
	intermediate	68	0.22	0.11	0.01	0.00	0.43	
	low	87	0.16	0.11	0.01	−0.06	0.38	
Fluency	high	150	0.52	0.21	0.93	2.47	0.01	0.43
	intermediate	68	0.40	0.23	0.85	1.74	0.08	
	low	186	0.57	0.13	0.83	4.37	0.00	
Total	high	150	0.37	0.26	0.48	6.56	0.00	3.73
	intermediate	68	0.16	−0.13	0.44	1.09	0.27	
	low	209	0.24	0.13	0.34	4.50	0.00	

Notes. CI = confidence interval; LL = lower limit; UL = upper limit. ***p* < .01.

of repetition type was not significant. Both ER and PR were found to improve the accuracy, lexical complexity, and fluency of L2 learners’ writing. Although PR negatively affected syntactic complexity in L2 learners’ writing, this effect was not statistically significant.

Table 7 illustrates that only syntactic complexity is moderated by language proficiency. For high-proficiency learners, TR could significantly improve the syntactic complexity of their writing with a small effect size. In contrast, for low-proficiency learners, syntactic complexity actually deteriorated.

Table 8 shows that previous studies on the effect of TR on L2 writing have primarily focused on adults. The data indicated that adult learners experienced improvements in all dimensions of CALF through TR. The total effect suggested that TR significantly enhanced the WP of adult learners, whereas no significant improvement was observed for YLs. Although the difference in effect size was not statistically significant, it can still be argued that TR is a more effective and beneficial tool for adult learners.

Table 9 demonstrates that writing genres can serve as an effective moderator between TR and L2 learners’ WP. A detailed analysis of different dimensions of CALF reveals significant differences in syntactic complexity and lexical complexity, with L2 learners achieving the greatest improvement in argumentative writing tasks. The effect sizes across all dimensions of CALF indicate that TR is most beneficial for argumentative writing. Lexical complexity improves in both argumentative and narrative writing, whereas in decision- making writing, the effect of TR is negative.

Table 10 reveals that previous studies on writing TR predominantly focus on individual writing tasks. No significant differences were identified across any dimensions of

Table 8. Moderating effect of the age of second language (L2) learners.

Sub-moderators		N	g	95% CI		Z	p	Q
				LL	UL			
Syntactic complexity	Adult	361	0.11	-0.10	0.33	1.02	0.31	2.00
	YL	36	-0.17	-0.49	0.15	-1.02	0.31	
Accuracy	Adult	361	0.27	0.07	0.01	0.13	0.41	0.01
	YL	36	0.30	0.31	0.09	-0.31	0.90	
Lexical complexity	Adult	224	0.29	0.06	0.00	0.17	0.41	2.18
	YL	57	0.05	0.15	0.02	-0.24	0.34	
Fluency	Adult	323	0.56	0.10	0.75	5.81	0.00	0.01
	YL	57	0.54	0.20	0.93	2.74	0.01	
Total	Adult	346	0.31	0.23	0.40	7.29	0.00	0.05
	YL	57	0.11	-0.08	0.30	1.13	0.26	

Notes. CI = confidence interval; LL = lower limit; UL = upper limit.

Table 9. Moderating effect of writing genres.

Sub-moderators		N	g	95% CI		Z	p	Q
				LL	UL			
SC	argumentative	150	0.40	0.29	0.51	7.11	0.00	20.57**
	decision making	30	-0.04	-0.21	0.14	-0.41	0.68	
	narrative	217	-0.17	-0.61	0.27	-0.76	0.45	
A	argumentative	150	0.36	0.11	0.01	0.14	0.58	0.65
	decision making	30	0.25	0.28	0.08	-0.30	0.81	
	narrative	217	0.25	0.08	0.01	0.09	0.41	
LC	argumentative	150	0.39	0.08	0.01	0.24	0.55	10.15**
	decision making	30	-0.06	0.12	0.01	-0.28	0.17	
	narrative	101	0.26	0.12	0.01	0.03	0.49	
F	argumentative	150	0.57	0.12	0.81	4.59	0.00	2.83
	decision making	30	0.79	0.15	1.08	5.35	0.00	
	narrative	200	0.43	0.16	0.74	2.74	0.01	
Total	argumentative	150	0.41	0.34	0.49	10.56	0.00	11.06**
	decision making	30	0.12	-0.08	0.32	1.14	0.25	
	narrative	223	0.26	0.17	0.35	5.85	0.00	

Notes. SL = syntactic complexity; LC = lexical complexity; A = accuracy; F = fluency; CI = confidence interval; LL = lower limit; UL = upper limit. ** $p < .01$.

CALF between individual and collaborative writing tasks. This finding suggests that L2 learners can benefit equally from TR, whether engaged in collaborative writing tasks or individual work, in improving their WP.

Table 11 indicates that the writing medium can serve as a mediator between TR and L2 learners' WP. The data show that computer-assisted writing repetition has a

Table 10. Moderating effect of interaction style.

	Sub-moderators	N	g	95% CI		Z	p	Q
				LL	UL			
Syntactic complexity	collaborative	87	−0.49	−1.50	0.52	−0.95	0.34	1.83
	individual	334	0.21	0.10	0.33	3.72	0.00	
Accuracy	collaborative	87	0.10	0.14	0.02	−0.18	0.38	0.94
	individual	334	0.27	0.09	0.01	0.09	0.44	
Lexical complexity	collaborative	20	0.41	0.42	0.18	−0.41	1.23	0.16
	individual	285	0.24	0.06	0.00	0.13	0.35	
Fluency	collaborative	49	0.59	0.18	0.94	3.34	0.00	0.20
	individual	355	0.50	0.11	0.72	4.39	0.00	
Total	collaborative	72	0.28	0.10	0.46	3.06	0.00	0.07
	individual	335	0.25	0.14	0.36	4.46	0.00	

Notes. CI = confidence interval; LL = lower limit; UL = upper limit.

Table 11. Moderating effect of writing medium.

	Sub-moderators	N	g	95% CI		Z	p	Q
				LL	UL			
Syntactic complexity	computer	160	0.30	0.15	0.45	4.03	0.00	5.50*
	pen-and-paper	261	−0.08	−0.37	0.20	−0.57	0.57	
Accuracy	computer	160	0.36	0.09	0.01	0.18	0.54	1.81
	pen-and-paper	261	0.17	0.11	0.01	−0.04	0.38	
Lexical complexity	computer	90	0.45	0.10	0.01	0.25	0.65	6.25*
	pen-and-paper	215	0.15	0.07	0.00	0.02	0.28	
Fluency	computer	160	0.50	0.14	0.77	3.57	0.00	0.00
	pen-and-paper	244	0.50	0.13	0.75	3.91	0.00	
Total	computer	160	0.38	0.29	0.47	8.55	0.00	5.72*
	pen-and-paper	267	0.19	0.07	0.32	2.98	0.00	

Notes. CI = confidence interval; LL = lower limit; UL = upper limit. * $p < .05$.

positive effect on all dimensions of CALF, with effect sizes generally exceeding those observed in pen-and-paper writing. This is especially evident in syntactic complexity and lexical complexity, where computer-assisted writing demonstrates significantly greater benefits.

V Discussion

I The effect of TR on L2 learners’ WP

This meta-analysis synthesized previous research on the effect of TR on L2 learners’ WP and revealed that TR could enhance L2 learners’ overall WP and all dimensions of CALF

although the effect varies. This finding supported the aforementioned theoretical frameworks, which suggest that as L2 learners become more familiar with the content or procedures of tasks, more attentional and processing capacity will be freed up, allowing them to allocate more focus to other dimensions of writing production and eventually gain an overall development.

A detailed examination of the different dimensions of CALF indicated that TR significantly improved L2 learners' writing fluency with a medium effect size ($g = 0.51, p < .01$), and accuracy and lexical complexity with an equally small effect size ($g = 0.24$ for both, $p < .01$ for both). While the positive effect on syntactic complexity was minimal and not statistically significant ($g = 0.06, p = .57$), the limited improvement in syntactic complexity may be attributed to several factors. First, L2 learners' attentional resources are limited. The additional attentional capacity released through TR may tend to be allocated to accuracy, lexical complexity, and fluency, making it more difficult to improve syntactic complexity. Second, learners at different proficiency levels may experience different challenges in enhancing syntactic complexity through repeated tasks. For beginner learners, the lack of sufficient syntactic knowledge may limit the potential for improvement (Y. Kim et al., 2020, 2022). While for more advanced learners, gains in lexical complexity may come at the cost of global syntactic complexity (Sánchez et al., 2020).

Previous studies often treated syntactic complexity and lexical complexity as a single dimension (e.g. Hidalgo & Lázaro-Ibarrola, 2020) or ignored lexical aspect (e.g. Amiryousefi, 2016; Y. Kim et al., 2020, 2022) when evaluating the complexity of L2 learners' writing. However, our analysis reveals the need to distinguish between the two measures.

According to Skehan (2009), the CAF framework was insufficient for evaluating L2 learners' performance. While complexity may be unidimensional for native speakers, it is not the case for non-native speakers. For non-native speakers, searching for more complex lexical choices demands additional attentional resources that may exert a trade-off effect on other dimensions, and a great probability is that syntactic complexity or accuracy is impaired. From our analysis, we could see clearly that accuracy prevailed in this battle, which is in line with Skehan's (2009) claim that 'tasks based on concrete or familiar information advantaged accuracy and fluency' (p. 511).

The differing impact of TR on syntactic complexity and lexical complexity may be attributed to L2 learners' greater reliance on their declarative system. According to Ullman's (2006) DP model, declarative systems support item-based aspects, such as the lexicon, while procedural systems govern rule-based aspects of language production, such as syntax. During TR, declarative systems are activated by previously familiar knowledge, resulting in greater improvements in lexical diversity, richness, and sophistication. In contrast, syntactic complexity relies more heavily on procedural systems and demands more cognitive processing, typically being achieved only after simpler aspects, such as lexical complexity and fluency, have become automatic.

2 The moderators between TR and L2 learners' WP

Considering the considerable heterogeneity of the effects, this study examined the moderating variables that influence the impact of TR on L2 learners' WP. Subgroup analysis

revealed that the TR interval, writing genre, and writing medium acted as effective moderators of the relationship between TR and L2 learners' WP, while repetition type, repetition number, the age and language proficiency of L2 learners, and interaction style did not. The varying impact of these moderators and sub-moderators underscored the complexity of TR.

Different TR intervals exhibited significantly different effect sizes. Our study indicated that a one-week interval is the most beneficial for L2 learners' WP, with positive effects observed across all dimensions of CALF. These findings align with previous meta-analyses suggesting that distributed practice is more advantageous than mass practice (e.g. S.K. Kim & Webb, 2022; Latimier et al., 2021). According to Mitchell et al. (2019), distributed practice helped embed representations in memory and gave these representations a higher 'resting level' of activation (p. 185). This facilitated easier retrieval of relevant knowledge and aided the transfer of declarative knowledge to procedural knowledge in repeated tasks. The cognitive capacity released during proceduralization enables L2 learners to access linguistic features more easily, enhancing their performance across CALF dimensions. However, our analysis also showed that as the gap between repetitions increased, the positive effect diminished, and when the gap reached four weeks, the effect turned negative. This may be attributed to memory decay, illustrating that the benefits of TR are time-constrained. Rogers (2023) conducted a conceptual review of the spacing effects of TR on L2 learners' oral production and found that several variables, such as the manipulation of post-tests, the number of repetitions, and the criteria for learning, can moderate the spacing effects. Given the scarcity of research specifically focusing on the spacing effects of TR on L2 learners' WP, further exploration is warranted to investigate the complex relationship between intervals and TR effects on L2 learners' WP.

Our study found that ER and PR could not moderate the effect of TR on all dimensions of CALF ($p > .05$). This finding contradicted some previous studies (e.g. Amiryousefi, 2016; Y. Kim et al., 2020), which argued that ER has differential effects on L2 learners' WP compared to PR, while aligning with others (e.g. Y. Kim & Tracy-Ventura, 2013). Our study confirmed that being familiar with a specific task procedure is as beneficial as with both the content and the procedure. According to Qiu and Lo (2017), this may be attributed to the engagement effect experienced by learners. L2 learners tended to be more motivated when repeating unfamiliar topics, whereas when repeating the same task, both their behavioral and cognitive engagement were impaired. This is also supported by Hidalgo and García Mayo's (2021) study, which found that the PR group produced more LREs than the ER group. The different content of the tasks may help maintain L2 learners' motivation and draw more attention to language form during writing production. Consequently, the impaired motivation and engagement in ER may offset the benefits derived from the attentional resources released due to previous experience with the content, resulting in a lack of significant difference between ER and PR.

Our meta-analysis revealed that the total effect of TR on the CALF was not moderated by proficiency ($g > .00, p > .05$), which aligns with Bygate's (2018) assertion that learners of all proficiency levels are likely to benefit from TR. However, a closer examination of the different dimensions of CALF indicated a distinct proficiency-related advantage

in syntactic complexity ($p = .006 < .01$) which is in line with Abdi Tabari and Golparvar's (2024) findings. Specifically, higher proficiency in a second language is associated with greater benefits from TR in terms of syntactic complexity. Conversely, for low-proficiency language learners, the effect of TR on syntactic complexity is even negative ($g = -0.12$, $p = .50$). This outcome may be attributed to the limited processing capacity of low-proficiency learners, which results in fewer attentional resources available for syntactic complexity as they prioritize fluency, accuracy, and lexical complexity. Additionally, low-proficiency language learners may lack complex syntactic knowledge, and the inherently demanding nature of syntactic complexity means that using multiple clauses in their writing may exceed their current capabilities (Y. Kim et al., 2020). Our findings diverged from Sánchez et al.'s (2020) study, which argued that lexical complexity, rather than syntactic complexity, is proficiency-dependent. This discrepancy may arise from two factors: first, most of the studies included in our analysis focused on narrative or argumentative writing tasks, while Sánchez et al.'s study concentrated on decision-making tasks, which may require varying levels of attention to different writing aspects. Second, Sánchez et al.'s study had a small sample size, with fewer than ten participants in each group, which may have influenced their conclusions.

Our study found that the writing genre served as an effective moderator between TR and L2 learners' WP, addressing questions raised by some researchers (e.g. López, 2018; Sánchez et al., 2020). The data indicated that L2 learners benefited most from argumentative writing tasks. Among the different dimensions of CALF, significant differences in syntactic and lexical complexity were observed across the three genres (i.e. narrative, argumentative, and decision making). According to Robinson's (2001) Cognition Hypothesis, cognitively demanding tasks are more likely to draw learners' attention to linguistic forms, thereby fostering more complex language use. Since argumentative writing requires greater cognitive reasoning, learners are more likely to make full use of their attentional resources to produce more complex language. Additionally, our study showed considerable improvement in lexical complexity during narrative TR. In contrast, for the decision-making writing task, TR did not lead to a more complex performance in terms of lexical complexity ($g = -0.06$), which aligns with findings of López (2018) and Sánchez et al. (2020). According to Skehan (2009), decision-making tasks were more negotiable and interactive. Learners may not utilize many less frequent words to pack their contribution. In contrast, narrative tasks, which are more input-driven and less negotiable, require the use of more difficult-to-avoid lexical items, as demonstrated in our study ($g = 0.26$). The characteristics of different genres dictate the allocation of attentional resources during TR.

Our study found that repetition number, interaction style, and learner's age did not serve as moderators between TR and any dimensions of CALF. However, these factors did reveal trends that may inform the implementation of TR and deepen our understanding of its effects. Regarding repetition number, our findings indicated that one repetition is sufficient to enhance L2 learners' WP, showing improvements in all dimensions of CALF ($g = 0.09, 0.21, 0.23, 0.52$), while when the repetition is added to five times, there is no significant improvement. In terms of cooperation style, while many studies have highlighted the benefits of collaborative writing, attributing part of the improvement to

interaction and negotiation between pairs (e.g. Hidalgo & García Mayo, 2021; Y. Kim et al., 2022), our meta-analysis did not find a significant difference between individual and collaborative tasks. Both individual TR and collaborative TR positively impacted students' WP. As for the age of learners, although no significant difference was observed in the effects of TR on adults versus YLs, effect sizes favored adults in terms of syntactic complexity ($g = 0.11$ vs. -0.17), lexical complexity ($g = 0.29$ vs. 0.05), while fluency ($g = 0.56$ vs. 0.54) and accuracy effect sizes were similar ($g = 0.27$ vs. 0.30). This suggests that TR may be a more effective tool for adults. According to Ullman's (2006) DP model, children rely more on procedural memory than adults, and as individuals age, they increasingly depend on the declarative system for processing vocabulary, morphology, and syntax. Since declarative knowledge, which is easier to transfer during TR, is rooted in declarative memory, adults may benefit more from knowledge transferred through previous practice, allowing them to free up more attentional resources to focus on language form. This reliance on declarative systems, which support item-based aspects, may explain why adults demonstrate greater improvements in lexical complexity in their writing. Furthermore, as adults are more cognitively mature, they possess more cognitive resources, enabling them to achieve a more balanced WP during repeated tasks.

In addition to the aforementioned moderators, this study found that the writing medium significantly influences L2 learners' WP. Specifically, computer-assisted writing repetition enhances all dimensions of CALF, with effect sizes generally greater than those observed in pen-and-paper writing, particularly in syntactic complexity and lexical complexity. The computer's associative input and auto-correction features can significantly alleviate the burden on fluency and accuracy, thereby freeing up attentional resources for generating more complex language. As digital devices become increasingly prevalent in SLA, these findings underscore the advantages of electronic tools in enhancing learners' WP.

VI Conclusions, limitations and implications for future studies

This study confirmed the effectiveness of TR on L2 writing, demonstrating an overall positive impact on L2 learners' WP. However, the development of different dimensions of CALF varied. L2 learners tend to prioritize improvements in writing fluency, accuracy, and lexical complexity over syntactic complexity in repetitive writing tasks. Three moderating factors, including TR intervals, writing genre, and writing medium, exhibited distinct effects. The findings indicate that TR is more effective when the interval is one week, particularly beneficial for argumentative writing tasks, and significantly enhanced when conducted in a computer-mediated environment. Although other factors, such as repetition number, repetition type, language proficiency and age of the learners, and interaction style did not moderate the effect of TR, they revealed general trends that provide valuable insights for implementing TR. For instance, it can be argued that a single repetition is sufficient to improve learners' WP and may serve as a more efficient tool for adult learners. Overall, TR proves to be an effective strategy for improving L2 learners' WP, regardless of their language proficiency, repetition type, or interaction style.

The findings offered valuable insights for Kellogg's (1996) writing model, Skehan's (2014) LAC hypothesis, SAT, Robinson's (2001) Cognition Hypothesis and Ullman's (2006) DP model, while also linking TR to conative factors such as learner engagement and motivation. Rather than reiterating sentences and phrases, TR serves as a process for transferring prior knowledge and creating new connections. The cognitive resources freed up during repetition allow learners to focus more effectively on language form. The examination of the effects of different moderators on TR's impact on L2 learners' WP provides teachers with insights to tailor TR activities to their students' characteristics and instructional content. For example, TR could be more effectively employed in argumentative writing tasks with a one-week interval. Generally, one repetition is often sufficient to facilitate TR and help alleviate learners' fatigue and boredom. While TR is a universally applicable approach regardless of learners' proficiency levels, L2 instructors should provide additional guidance for cognitively demanding processes, such as generating complex sentences, especially for low-proficiency learners, to help them alleviate cognitive burdens.

This study has several limitations. First, this study relied on the framework of CALF as the sole criterion to evaluate students' WP. While CALF is generally considered a reliable tool for measuring writing proficiency, its effectiveness has faced scrutiny in recent years. For instance, Hidalgo and Lázaro-Ibarrola (2020) employed both CAF and holistic ratings to examine the effect of TR on collaborative writing of YLs and found that CAF measures could not fully capture improvements in learners' WP, whereas holistic measures could. This discrepancy may arise from factors such as writing length and sample size (Hidalgo & Lázaro-Ibarrola, 2020), suggesting that a combination of both types of measurements could offer a more comprehensive assessment. Second, as noted by Hiver et al. (2024), group-level patterns may not always apply to each individual, and the same applies in reverse. Each learner exhibits a unique developmental trajectory in CALF measures, which may not align with overall group trends. Therefore, it is important for researchers to pay attention to individual differences and personalized development in future studies. Third, this study only examined the moderating effects of TR on L2 learners' WP, overlooking potential interactions between different factors. Future research could adopt a cross-analysis approach to explore how combined moderators of TR influence L2 learners' WP.

Acknowledgements

We would like to express our sincere gratitude to Prof. Luo and the anonymous reviewers of the manuscript.

Author contributions

Junge Liu (first author) conceptualized the study, performed data analysis, and drafted the initial manuscript. Yinjie Tang (second author) was responsible for revising the manuscript.

Data availability

The data supporting the findings of this study are available upon request from the corresponding author.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The present study was supported by the project ‘School Age Cohort Study of Brain and Mind Development in China’ (STI 2030 – Major Projects 2021ZD0200500).

Ethical considerations

The study was conducted in accordance with ethical guidelines and complied with ethical requirements.

Consent to participate

Informed consent was obtained from all participants involved in the study.

Consent for publication

All participants provided consent for the publication of the results from this study

ORCID iDs

Junge Liu  <https://orcid.org/0009-0006-0592-5741>

Yinjie Tang  <https://orcid.org/0009-0004-5742-2981>

Notes

1. The third repetition.
2. The second repetition.

References

Note. References with an asterisk indicate studies that were included in the meta-analysis.

- *Abdi Tabari, M., & Golparvar, S.E. (2024). The interplay of task repetition and task rehearsal in L2 written production across varied proficiency levels. *Language Teaching Research*. Advance online publication. <https://doi.org/10.1177/13621688241266940>
- Abdi Tabari, M., Johnson, M.D., & Farahanynia, M. (2023). Automated analysis of cohesive features in L2 writing: Examining effects of task complexity and task repetition. *Assessing Writing*, 58, Article 100783.
- *Abdi Tabari, M., Khezlrou, S., & Ghanbar, H. (2024). Task repetition versus task rehearsal: Understanding effects of task-readiness factors and elemental genres on L2 writing task performance. *Language Teaching Research*. Advance online publication. <https://doi.org/10.1177/13621688241249689>
- Abdi Tabari, M., Khezlrou, S., & Tian, Y. (2024). Verb argument construction complexity indices and L2 written production: Effects of task complexity and task repetition. *Innovation in Language Learning and Teaching*, 18, 1–16.
- Ahmadian, M.J. (2013). Working memory and task repetition in second language oral production. *Asian Journal of English Language Teaching*, 23, 37–55.
- *Amiryousefi, M. (2016). The differential effects of two types of task repetition on the complexity, accuracy, and fluency in computer-mediated L2 written production: A focus on computer anxiety. *Computer Assisted Language Learning*, 29, 1052–1068.
- Borenstein, M. (2009). *Introduction to meta-analysis*. John Wiley.

- Bui, G., & Yu, R. (2021). Differentiating task repetition from task rehearsal. In N.P. Sudharshana & L. Mukhopadhyay (Eds.), *Task-based language teaching and assessment* (pp. 119–137). Springer Nature Singapore.
- Bygate, M. (2001). Effects of task repetition on the structure and control of oral language. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks* (pp. 23–48). Longman.
- Bygate, M. (Ed.). (2018). *Learning language through task repetition*. John Benjamins.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Psychology Press.
- DeKeyser, R. (2018). Task repetition for language learning: A perspective from skill acquisition theory. In M. Bygate (Ed.), *Learning language through task repetition* (pp. 27–41). John Benjamins.
- Durlak, J.A. (2009). How to select, calculate, and interpret effect sizes. *Journal of Pediatric Psychology*, 34, 917–928.
- Ellis, R. (2005). Planning and task-based performance: Theory and research. In R. Ellis (Ed.), *Language learning and language teaching: Volume 11* (pp. 3–34). John Benjamins.
- Ellis, R., Skehan, P., Li, S., Shintani, N., & Lambert, C. (2020). *Task-based language teaching: Theory and practice*. Cambridge University Press.
- Fukunaga, T. (2023). L2 writing development through two types of writing task repetition. *International Review of Applied Linguistics in Language Teaching*, 61, 1109–1138.
- Gass, S., Mackey, A., Alvarez-Torres, M.J., & Fernández-García, M. (1999). The effects of task repetition on linguistic output. *Language Learning*, 49, 549–581.
- Hanzawa, K., & Suzuki, Y. (2023). How do learners perceive task repetition? Distributed practice effects on engagement and metacognitive judgment. *Modern Language Journal*, 107, 451–478.
- Hidalgo, M.Á., & García Mayo, M.D.P. (2021). The influence of task repetition type on young EFL learners' attention to form. *Language Teaching Research*, 25, 565–586.
- *Hidalgo, M.Á., & Lázaro-Ibarrola, A. (2020). Task repetition and collaborative writing by EFL children: Beyond CAF measures. *Studies in Second Language Learning and Teaching*, 10, 501–522.
- Hiver, P., Al-Hoorie, A.H., & Murakami, A. (2024). Modeling the effects of task repetition in second language writing: Examining interindividual and intraindividual variability. *Language Learning*, 75, 493–532.
- Housen, A., Kuiken, F., & Vedder, I. (Eds.). (2012). *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA: Volume 32*. John Benjamins.
- Huh, M.-H., Lee, J., & Ha, H.S. (2018). EFL writing development through repetition of a literature-reading-writing task. *Korean Journal of Applied Linguistics*, 34, 53–74.
- Hunter, A.M. (2017). Fluency development in the ESL classroom: The impact of immediate task repetition and procedural repetition on learners' oral fluency. Unpublished doctoral dissertation, St Mary's University, Twickenham, London, UK.
- Itze Arredondo-Tapia, D., & Emmanuell Garcia-Ponce, E. (2021). Role of task repetition and content familiarity in EFL students' fluency and accuracy in narrative tasks: A case study. *Journal of Language and Education*, 7, 45–63.
- Jung, S. (2013). The effect of task repetition and corrective feedback in L2 writing. *MSU Working Papers in Second Language Studies*, 4, 24–38.
- Kellogg, R.T. (1996). A model of working memory in writing. In M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences and applications* (pp. 57–71). Erlbaum.
- Khezlrou, S. (2021). Effects of timing and availability of isolated FFI on learners' written accuracy and fluency through task repetition. *Language Learning Journal*, 49, 568–580.

- Kim, J.C. (2024). The role of individual differences and effect of task repetition with written corrective feedback in L2 writing development. Unpublished doctoral dissertation, Florida State University, Tallahassee, FL, USA.
- Kim, J.K., & Lee, J.H. (2019). Effects of the number of task repetitions on Korean middle school students' English writing and perception. *Secondary English Education*, 12, 3–25.
- Kim, S.K., & Webb, S. (2022). The effects of spaced practice on second language learning: A meta-analysis. *Language Learning*, 72, 269–319.
- *Kim, Y., Choi, B., Yun, H., Kim, B., & Choi, S. (2022). Task repetition, synchronous written corrective feedback and the learning of Korean grammar: A classroom-based study. *Language Teaching Research*, 26, 1106–1132.
- *Kim, Y., Kang, S., Yun, H., Kim, B., & Choi, B. (2020). The role of task repetition in a Korean as a foreign language classroom: Writing quality, attention to form, and learning of Korean grammar. *Foreign Language Annals*, 53, 827–849.
- Kim, Y., & Tracy-Ventura, N. (2013). The role of task repetition in L2 performance development: What needs to be repeated during task-based interaction? *System*, 41, 829–840.
- Lambert, C., Kormos, J., & Minn, D. (2017). Task repetition and second language speech processing. *Studies in Second Language Acquisition*, 39, 167–196.
- Latimier, A., Peyre, H., & Ramus, F. (2021). A meta-analytic review of the benefit of spacing out retrieval practice episodes on retention. *Educational Psychology Review*, 33, 959–987.
- Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning*, 60, 309–365.
- Lipsey, M.W., & Wilson, D.B. (2001). *Practical meta-analysis*. SAGE Publications.
- Liu, J. (2024). Meaning-oriented understanding of L2 academic writing development with task-type repetition. *Humanities and Social Sciences Communications*, 11, Article 1167.
- López, A.J.S. (2018). External task-repetition: The role of modality, written corrective feedback and proficiency. A comparative study. Unpublished doctoral dissertation, Universidad de Murcia, Murcia, Spain.
- Luo, S.Q., & Xun, H. (2023). Task-based language learning research at home and abroad (2001–2021): Current status, trends and new frontiers. *Shandong Foreign Language Teaching*, 44, 39–48.
- *Mehrang, F. (2016). The effect of task structure, task repetition, and reformulation on foreign language written performance. Unpublished doctoral dissertation, University of Auckland, Auckland, NSW, Australia.
- Mitchell, R., Myles, F., & Marsden, E. (2019). *Second language learning theories* (4th ed.). Routledge.
- Nguyen, H., Nguyen, H., Vo, N., & Huynh, N. (2023). The combined effects of task repetition and post-task teacher-corrected transcribing on complexity, accuracy and fluency of L2 oral performance. *International Review of Applied Linguistics in Language Teaching*, 63, 1317–1345.
- Nitta, R., & Baba, K. (2014). Task repetition and L2 writing development: A longitudinal study from a dynamic systems perspective. In H. Byrnes & R. Manchón (Eds.), *Task-based language learning and teaching: Insights from writing* (pp. 107–136). John Benjamins.
- Patanasorn, C. (2010). Effects of procedural content and task repetition on accuracy and fluency in an EFL context. Unpublished doctoral dissertation, Northern Arizona University, Flagstaff, AZ, USA.
- Qiu, X., & Lo, Y.Y. (2017). Content familiarity, task repetition and Chinese EFL learners' engagement in second language use. *Language Teaching Research*, 21, 681–698.

- Robinson, P. (2001). Task complexity, cognitive resources, and syllabus design: A triadic framework for examining task influences on SLA. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 287–318). Cambridge University Press.
- Róg, T., & Krawiec, M. (2024). Task repetition and L2 grit: Exploring the interplay in language learners' oral performance. *TASK*, 4, 241–266.
- Rogers, J. (2023). Spacing effects in task repetition research. *Language Learning*, 73, 445–474.
- *Roothoof, H., Lázaro-Ibarrola, A., & Bulté, B. (2022). Task repetition and corrective feedback via models and direct corrections among young EFL writers: Draft quality and task motivation. *Language Teaching Research*, 29, 1311–1344.
- *Sánchez, A.J., Manchón, R.M., & Gilabert, R. (2020). The effects of task repetition across modalities and proficiency levels. In R. Manchon (Ed.), *Writing and language learning: Advancing research agendas* (pp. 121–144). John Benjamins
- Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, 30, 510–532.
- Skehan, P. (2014). Introduction: Coordinated research into task-based performance. In P. Skehan (Ed.), *Processing perspectives on task performance* (pp. 1–26). John Benjamins.
- Sun, Y., & Liu, H. (2023). A meta-analysis of the effectiveness of task repetition on L2 oral output. *Foreign Languages Research*, 40, 68–74.
- Thai, C., & Boers, F. (2016). Repeating a monologue under increasing time pressure: Effects on fluency, complexity, and accuracy. *TESOL Quarterly*, 50, 369–393.
- Ullman, M. (2006). The declarative/procedural model and the shallow structure hypothesis. *Applied Psycholinguistics*, 27, 97–105.