System 103 (2021) 102654

Contents lists available at ScienceDirect

System

journal homepage: www.elsevier.com/locate/system

A comparative study of the effects of L1 and L2 prewriting discussions on L2 writing performance

Jianmin Gao, Shangchao Min *

Institute of Applied Linguistics, Zhejiang University, Hangzhou, 310058, China

ARTICLE INFO

Keywords:
Collaborative planning
Prewriting discussion
L2 writing

ABSTRACT

Despite the frequent implementation of prewriting discussion in L2 writing classrooms, little attention has been paid to the issue that which language, L1 or L2, should be used for discussion. To advance our knowledge on this issue, we compared the effects of L1 and L2 prewriting discussions on L2 writing performance. We randomly assigned 66 university students to two groups in terms of the language they were required to use throughout the discussion: the L1 group and the L2 group. Within each group, they formed pairs to talk about a given writing topic, after which they individually finished their writing and completed an online retrospective questionnaire. We compared the writings of the two groups in terms of complexity, accuracy, fluency, content richness, and holistic quality. The results suggested that: 1) the L1 group outperformed the L2 group in lexical complexity, accuracy, and fluency, while there were no between-group differences in other measures; 2) the L1 group overall held more positive views towards the effect of the discussion, especially in terms of the facilitative role of discussion in improving fluency. These findings point to the need of a balanced use of students' L1 and L2 in L2 writing classrooms.

1. Introduction

With more attention given to task-based language teaching (TBLT) in the field of L2 education, many researchers have made efforts to investigate the factors that impact learners' writing task performance, including task type, task complexity, learner characteristics, and prewriting planning (e.g., Adams et al., 2015; Azimzadeh, 2017; Cho, 2018; Frear & Bitchener, 2015; Johnson, 2017; Michel et al., 2019; Pae, 2018; Rahimi & Zhang, 2019; Révész et al., 2017; Ruiz-Funes, 2015). According to Ellis and Yuan (2004), prewriting planning is a stage prior to writing that allows writers a certain amount of time to formulate plans, develop ideas, and retrieve the information needed. Essentially, there are two forms of prewriting planning: individual and collaborative planning. Prewriting discussion is a form of collaborative planning. Supported by multiple theoretical insights, such as sociocultural and interactionist theories, prewriting discussion has been frequently used in L2 writing classrooms (Liao, 2018; Neumann & McDonough, 2015). However, in authentic teaching scenarios, there are not yet clear guidelines on which language (L1 or L2) should be used for classroom discourse. This could be problematic since different languages used in prewriting discussion may produce different effects on writing quality; ignoring these potential differences can lead to lowered efficiency in language teaching and learning. As claimed in many studies on the L2 writing process, the use of L1 could improve L2 written outcomes (e.g., DiCamilla & Antón, 2012; Wang & Wen, 2002; Zhang, 2018). However, few empirical studies have investigated the role of L1 use in the planning stage. We attempted to address this issue by

* Corresponding author.
E-mail addresses: jimmy_gao@zju.edu.cn (J. Gao), msc@zju.edu.cn (S. Min).

https://doi.org/10.1016/j.system.2021.102654
Received 27 May 2020; Received in revised form 22 September 2021; Accepted 25 September 2021
Available online 2 October 2021
0346-251X/

comparing the effects of L1 and L2 prewriting discussions on L2 writing performance, as measured by various linguistic features and human ratings. Exploring the effects of L1 and L2 use in prewriting discussion may deepen our understanding of the role of L1 use in the prewriting stage, and offer insights into ways to improve teaching practice in L2 writing classrooms.

2. Literature review

2.1. Learner discussion in L2 writing

The use of learner discussion in L2 writing classes is primarily supported by Vygotsky's sociocultural theory of mind (Vygotsky, 1978). The activities external to learners are the sources of their mental activities or cognitive functions, such as memory and attention; language serves as a mediating tool for internalising external activities to psychological resources. Swain (1997, 2000) first linked collaborative dialogue to the sociocultural perspective of learning. She defined collaborative dialogue as a 'problem-solving, and hence knowledge-building dialogue' (Swain, 2000, p. 113). In the setting of L2 learning, Swain argued that dialogue—where peer learners verbalise their thoughts—mediates the process of building knowledge, and co-built knowledge, as a product of the dialogue, might become a resource pool for learners' individual use of their L2. The concept of collaborative dialogue has great pedagogical implications in promoting learners' engagement in collaborative work, and many researchers in L2 writing are dedicated to investigating the role of peer discussion at various stages before, throughout, and after writing.

Many studies have focused on the types of discussions that occur when learners cooperate to co-construct a text (e.g., DiCamilla & Antón, 2012; Storch, 2005; Storch & Aldosari, 2010; Wigglesworth & Storch, 2009; Zhang, 2018), or at the peer review stage after they individually complete their writing (e.g., Ho & Savignon, 2007; Lockhart & Ng, 1995; Nelson & Carson, 1998; Zaccaron & Xhafaj, 2020). Although findings indicate that these two types of discussions in L2 writing classrooms have many advantages, some researchers have argued that they both have drawbacks that inhibit their application when teaching L2 writing (Neumann & McDonough, 2015; Storch, 2005). For instance, Neumann and McDonough (2015), drawing upon earlier studies (Kagan, 1995; Storch, 2005; Strauss & U, 2007), summarised two practical concerns that language instructors might have when they employ collaborative writing activities in which learners discuss throughout the process of co-constructed writing: assessment and time constraints. Group assessment may threaten assessment fairness since it is difficult for instructors to evaluate individual efforts in a co-constructed text; therefore, grades on a collaboratively written text can barely capture individual writing ability. In addition, collaborative writing usually takes much longer, which is generally not feasible in tertiary-level L2 classrooms where time is usually limited. Regarding peer discussion at the review stage or conference peer feedback (Zaccaron & Xhafaj, 2020), Storch (2005) asserted that it is so product-oriented that learners cannot benefit from each other in the writing process. Compared with these types of discussions in L2 writing, prewriting discussion has its unique strengths; it allows learners to predict and jointly solve problems they may encounter in the writing process. However, only a few studies have explored collaborative prewriting discussion in L2 classrooms.

In the limited literature on prewriting discussion, most studies are comparative in nature, comparing the effects of individual and paired prewriting planning (Kang & Lee, 2019; McDonough & De Vleeschauwer, 2019), structured small-group prewriting discussion and individual planning (Li & Zhang, 2021), face-to-face prewriting discussion and online text chat (Kessler et al., 2020; Liao, 2018), or instructor-led, peer-led and no discussions (Shi, 1998). These studies suggest that prewriting discussion is helpful in improving writing fluency, accuracy and complexity. The participants in most of these studies were English L2 learners, except for the investigations by Liao (2018) and Kessler et al. (2020), in which the participants were studying Chinese. One underlying issue that might be problematic in these studies is that most of them did not clarify which language (i.e., L1 or L2) was used in the prewriting discussion, although it could be inferred that learners spoke their L2 during the discussion. As claimed by Aubrey and Leeming (2020), one essential factor that could influence the effect of planning is whether the discussion is carried out in the students' L1 or L2. Taking the use of L2 for granted in prewriting discussion while ignoring L1 use might limit the explanatory force of these findings.

2.2. L1 use in L2 writing

Many studies exploring L1 use in L2 settings have maintained that the use of L1 may contribute to L2 development. For example, Thoms et al. (2005) and Al Masaeed (2016) suggested that the proper use of learners' L1 could smoothen the process of doing a language task (e.g., a jigsaw or role-playing task) in L2 settings and promote L2 learning. Further, the balanced use of L1 and L2 in classrooms may be useful to sustain an active interaction between peers or between learners and instructors (Storch & Wigglesworth, 2003; Tsagari & Giannikas, 2018), and intermediate-level L2 learners could collaborate to solve form-focused problems more successfully when they are required to use their L1 in the process of collaboration (Scott & de la Fuente, 2008). Gánem-Gutiérrez and Roehr (2011) also implied that L1 is an important tool to solve problems when completing a form-focused task. This was echoed by a more recent study by Li et al. (2020). They explored the nature of 48 English major students' small-group talks before individual writing, finding that the students were inclined to use L1 when they attempted to clarify information and negotiate about the meaning of vocabulary. L2 learners should be given more opportunities to use their L2 in the classroom according to the communicative approach to language teaching (Storch & Wigglesworth, 2003); however, the studies above indicate that L1 use should not be ignored since it also plays an important role in L2 development.

Earlier L2 writing research has also shown that L2 writers tend to use their L1 during the writing process. For instance, in a case study by Lay (1982)—one of the earliest attempts at exploring L1 use in L2 writing—the think-aloud data of four Chinese English as a foreign language (EFL) university students showed that they relied on their L1 to organize their ideas during the writing process. Additionally, in a case study by Cumming (1990), 23 EFL adult learners' verbal reports demonstrated that when they could not recall

an appropriate item in their L2, they frequently turned to their L1 for the equivalent expression; they then verified the intended meaning before translating it into their L2. Cumming also affirmed that equating words and phrases in two languages is a process of language integration in which learners might integrate previously acquired L2 knowledge with their L1 knowledge in the writing process, and consolidate their L2 use. Moreover, in a study by Roca et al. (1999), the extensive use of L1 by five Spanish EFL learners was revealed when they expanded, elaborated, and restructured their ideas when writing in their L2. Later, more studies attempted to determine the extent to which one's L1 is used in the L2 writing process, either by analysing L2 writers' self-reports (Cohen & Brooks-Carson, 2001), or by exploring their think-aloud data (Guo & Liu, 1997; Manchon et al., 2000). Some researchers have taken further steps to investigate the interplay between the function and amount of L1 use and other individual or task factors, specifically whether the use of L1 varies across L2 proficiency (DiCamilla & Antón, 2012; Storch & Aldosari, 2010; Wang & Wen, 2002; Woodall, 2002), general writing proficiency (van Weijen et al., 2009), and task types (Storch & Wigglesworth, 2003). All of these studies more or less imply that learners do use their L1 to fulfil different goals in the L2 writing process. Particularly, a recent study by Wei et al. (2020) found that Chinese EFL learners' L1-to-L2 rhetorical transfer in L2 argumentative writing was positively associated with their perceptions of L2 writing difficulty, which indicates that L1 rhetorical knowledge is a problem-solving asset for L2 writers to alleviate their sense of insecurity and promote writing coherence. However, most studies have focused on determining the nature of L1 use or language switching, ignoring the fact that the different choices of L1 and L2 use in classroom activities may produce different effects on L2 learning; almost none of them have examined the potentially differential effects on L2 writing performance, except for Zhang (2018) and Pu (2009).

Zhang (2018) focused on the use of L1 and L2 in collaborative writing, and examined the in-class argumentative essays of 35 pairs of Chinese intermediate EFL learners in terms of both linguistic measures and overall ratings to compare the effects of using different languages in the discussion throughout the co-construction process. L1 use during collaborative discussion contributes to higher syntactic complexity (measured by the mean length of clauses) and has the potential to reduce linguistic errors, but using different languages does not lead to significant differences in other aspects. In light of other studies that have investigated the role of L1 use in collaborative writing from a sociocultural standpoint (e.g., Swain & Lapkin, 2000), Zhang argued that L1 might serve as a meta-linguistic tool to promote form-focused negotiation and easier retrieval of complex sentence structures in L2. Zhang's study greatly contributed to the research progress of L1 and L2 use in collaborative writing; however, less is known about the role of L1 and L2 in prewriting discussion.

A noteworthy study was conducted by Pu (2009), who explored the effects of prewriting discussion in different languages (L1 versus L2) on the language quality of individual writings; Pu reported that the L2 group did better than the L1 group in complexity, measured by the ratio of clauses to T-units. However, there were only six participants in each group, and they were all Chinese students majoring in English, which might limit the generalisability of Pu's findings. In addition, Pu only included one measure of complexity, with no discrimination between lexical and syntactic complexity; other dimensions of composition such as content, organisation, and holistic quality were not considered. Both Pu and Zhang's studies, as well as many other studies on prewriting discussion, did not survey learners' perceptions of peer discussion in different languages. Such perceptions may be reflective of the task effect on students' motivations for learning L2 writing, which is of vital importance for evaluating teaching tasks.

In sum, compared to collaborative writing discussion and conference peer feedback, prewriting discussion is a more flexible, process-oriented activity that can be applied in L2 writing classrooms. However, little research has been conducted to investigate the practical effects of prewriting discussion. Further, although many studies suggest that L1 use in L2 writing has certain benefits for improving writing quality, studies comparing the effects of L1 and L2 use in prewriting discussion are quite rare. Although Pu (2009) made an earlier attempt, the findings might not be generalisable due to research limitations. To fill the knowledge gap and offer pedagogical implications for teaching L2 writing, using a mixed method, we compared the effects of L1 and L2 prewriting discussions on Chinese English learners' writing performance. We formulated two research questions:

1 Do prewriting discussions in L1 and L2 have different effects on text quality in terms of complexity, accuracy, fluency (CAF), content richness, and holistic performance?
2 How do the participants perceive the effects of L1 and L2 prewriting discussions?

3. Methodology

3.1. Participants

Sixty-six EFL learners (15 males, 51 females) from the same university in China volunteered to participate. They were freshmen with different majors, and were from two parallel classes of an English course taught by the same instructor at the time of the study. Their mean age was 18.3 years ($SD = 0.61$), and the average time they had studied English for was 10.24 years ($SD = 1.07$). They had learned English mainly in the classroom; none of them had the experience of being abroad. Thus, the participants shared a similar English writing experience. All participants had received regular trainings on collaborative discussion in class prior to our experiment because the instructor often employed collaborative planning as a teaching activity in class. The university considered the participants to be English learners at the intermediate level based on their performance on a placement test composed of four parts: (1) listening comprehension, (2) vocabulary, (3) grammar, and (4) reading comprehension. Based on the score percentile on the placement test, all freshmen were assigned to either an advanced or intermediate class for a tailored college English course at the beginning of the semester. The participants were all from the intermediate class; according to the university's English department, students placed into this class are at around the B1 level of the Common European Framework of Reference (CEFR).

We directly treated two classes of participants as two groups: the L1 group included 18 pairs (5 males, 31 females) that held prewriting discussions in their L1; the L2 group comprised 15 pairs (10 males, 20 females) that conducted prewriting discussions in their L2. We gathered the participants' writing scores on the mid-term English test prior to the experiment to gauge their writing proficiency. An independent writing task, with a maximum of 15 points to be earned, was used in the exam; the students were required to write an essay of 120 words or more on the topic 'Whether English classes should be compulsory for college students'. The instructor rated all the essays using the writing criteria of the College English Test Band 4 (CET-4), which is a national English test for non-English majors. For the writing performance of the two groups on the mid-term exam, the L1 group had a mean score of 10.28 ($SD = 1.32$), whereas the L2 group had a mean score of 10.17 ($SD = 1.12$). Since their scores were not normally distributed, we used a Mann–Whitney U test to compare the difference between the two groups. There was no significant difference in their mean scores ($U = 481$, $z = -0.786$, $p = .432$). Therefore, we inferred that the two groups were homogeneous in terms of writing proficiency.

In addition to the student participants, two raters including the course instructor and the first author who have the experience of scoring English writings of Chinese college students participated in the study.

3.2. Data collection

The participants were told to form pairs on their own for their prewriting discussion two days before it started. The instructor and the first author collected the data, including the recordings of the discussions and the post-discussion writings within one day. The data of the L1 and L2 groups were collected at the same time in two different multimedia classrooms. Each participant was required to wear the headphones and click the 'start' button on the machine when they began to talk with their partners so that the entire process of their interaction could be automatically recorded; only the two students with adjacent seat numbers could hear each other's voice through their headphones (e.g., No. 1 and No. 2; No. 3 and No. 4). For each group, the participants were asked to have a discussion within 10 min after they read the prompt of the independent writing task (see Appendix A). An independent writing task, rather than an integrated one, was used since most participants were not familiar with the latter, and such unfamiliarity had the potential to influence the results. The genre of argumentative writing was selected because of its widespread use in the language tests for Chinese EFL learners (Huang & Zhang, 2020). The first author devised the topic of the writing task with help from the instructor. The writing topic of being a smart consumer was related to the theme of one chapter in the students' textbook and was familiar to everyone, such that there would not be any topic bias. After they finished their discussion, they could start their individual writings. The time limit was 40 min, but no limit was set on the number of words the participants could write to yield a better measurement of their writing fluency. The time limits for discussion and writing were set based on a pilot study in which two pairs of students with similar profiles to the participants held a discussion on a given topic in English and then wrote an essay. All the task settings for the two groups were the same except for the language the participants were required to use in their discussion.

After finishing the writing task, the participants had to respond to an online questionnaire (see Appendix B) that consisted of five closed-ended questions and one open-ended question. The five closed-ended questions asked the participants about their views on the discussion's role in improving writing CAF, content richness, and overall quality on a 5-point Likert scale. In addition, they were asked to report their perceptions about the role of their discussion in an open-ended question at the end. To ensure that all participants understood the terms used in the questionnaire, the instructor and the researcher provided explanations before handing out the questionnaire; all questions were written in Chinese, the participants' native language.

Since one student's writing from the L1 group was too lengthy (almost three times as long as the others') and full of errors—thus having the potential to influence the final outcome—we excluded it from the subsequent analysis. Additionally, to increase the reliability of the results, we excluded the data from the questionnaire in which the participants assigned the same score across items. Eventually, we examined 65 essays and 51 participants' responses to the questionnaire.

3.3. Data analysis

We primarily analysed the participants' writings in terms of CAF, which has been widely employed in quantifying the effects of task complexity and planning on EFL learners' oral and writing performance (Armstrong, 2010; Johnson et al., 2012). Since the importance of assessing the content quality of writing and speaking performance has been well recognized (Cheung, 2016; Frost et al., 2011; Liao, 2018; Muhammad, 2019), we also examined the participants' writings in terms of content richness and holistic quality to better understand the difference between the effects of L1 and L2 use during the prewriting discussion. We used independent samples t-tests to detect potential differences in writing performance between the two groups. Prior to all the parametric tests, our inspection of Q-Q plots revealed that the data were approximately normally distributed for both groups. It should be noted that the p values were not adjusted in the subsequent report based on the concerns of some researchers (e.g., Althouse, 2016; Feise, 2002; Streiner, 2015). According to Streiner (2015), in a situation where group differences among the outcome variables are to be examined, the adjustment of the p values is not strictly necessary, especially if the outcomes are small in number and have been specified a priori. Since the current study served as an exploratory investigation, the p values were not adjusted. However, it is expected that the subsequent confirmatory studies with pre-planned hypotheses could be done to test the findings in this study.

Complexity consists of syntactic and lexical complexity. We analysed these two linguistic features of the participants' writings with the Coh-Metrix3.0 program (Graesser et al., 2004). It captures linguistic and organizational features of texts by applying techniques connected with cognitive psychology and computational linguistics. The usefulness of the Coh-Metrix indices in capturing the textual nuance has been demonstrated in more than 50 published studies (McNamara et al., 2010). Although Coh-Metrix was initially designed to quantify and evaluate textual cohesion, it also provides users with a variety of indices that can capture the characteristics of

syntactic complexity from multiple dimensionalities (McNamara et al., 2014). Crossley and McNamara (2014) used measures in Coh-Metrix to trace university-level EFL learners' growth in syntactic complexity; they found that traditional measures based on T-units could not consistently manifest L2 learners' syntactic development. In addition, Lu (2017) claimed that Coh-Metrix is one of the three tools commonly used to automatically analyse syntactic complexity, and that many indices of syntactic complexity provided by Coh-Metrix are good predictors of holistic writing quality. Therefore, due to its capability to afford speed, reliability and easy availability, we employed the Coh-Metrix3.0 program to scrutinise syntactic complexity. All indices related to syntactic complexity were included based on previous studies and descriptions of the indices provided by the program's developers. To select the indices most related to this study, we conducted a factor analysis. The dimensionality of the indices of syntactic complexity was analysed using principal component factor analysis. Three criteria were used to decide the number of the factors to retain: the eigenvalues-greater-than-1 criterion, the scree test, and the rotated factor solution by a varimax rotation procedure. In the end, three factors were extracted out of the six indices (see Table 1). The phrasal level factor accounted for 31.3% of the variance of the indices, the sentence level factor accounted for 30.2%, and the factor of noun phrase use accounted for 21.1%. For lexical complexity, we chose the measure of textual lexical diversity (MTLD)—which refers to 'the mean length of word strings that maintain a criterion level of lexical variation' (McCarthy & Jarvis, 2010, p. 381)—as the index since it is the least susceptible to text length (McCarthy & Jarvis, 2010).

Human coders gauged accuracy grounded in the definitions of the two types of errors proposed by Bardovi-Harling and Bofman (1989), and applied by Neumann (2014) and Plakans et al. (2016). According to Bardovi-Harling and Bofman (1989, p. 21), morphological errors refer to 'errors in nominal morphology (plural, number agreement, and possessive), errors in verbal morphology (tense, subject-verb agreement, and passive formation), errors in determiners and articles, and errors in prepositions'; syntactic errors are 'word order errors, errors resulting from absence of major or minor constituents, and errors in combining sentences'. They also proposed the type of lexical-idiomatic errors with an example; since they did not define such errors clearly, and idiomatic use is not frequent in the writings of intermediate English learners, only syntactic and morphological errors were coded. However, during our first pilot coding, these two kinds of errors could not cover all errors in the participants' writings. We then supplemented collocation errors, such as errors in combining verbs and nouns (e.g., learn knowledge) or adjectives and nouns (e.g., too much things). We counted—but did not classify—all of these errors in the analysis, as the issue on error types is out of the scope of this study. We randomly chose 13 writings for the second pilot coding of linguistic errors; there was a high agreement indicated by the Pearson coefficient ($r = .839$) between coders. After that, the first author coded the remaining writings. The raw frequency of errors might be affected by the length of the writing, which means that more words may lead to more errors. To mitigate this issue, we employed a ratio of accuracy (error counts/total word counts $\times$ 10), originally used by Plakans et al. (2016).

We gauged fluency in word counts, the same as that in the work of Pu (2009) and Zhang (2018). Most studies have merely adopted a rating scale to assess content richness (e.g., Kessler et al., 2020; Liao, 2018; McDonough & De Vleeschauwer, 2019; Shi, 1998; Zhang, 2018). This may lead to the incomparability of different findings, since the use of scoring rubrics varies across studies. Frost et al. (2011) innovatively devised an approach to establish the content quality of university-level EFL learners' speaking performance on an integrated listening-speaking task. They started from the input listening materials to generalise the list of key points, during which they discussed and reached a final consensus; they then tallied the number of key points included in the learners' oral performance as the indicator of content richness. The main point of this approach is to predetermine the criteria of the key idea units that are relevant to the writing or speaking topic, or consistent with the input materials if the task is an integrated one. With reference to the 'key points approach', we analysed the writing prompt, read through a random sample consisting of 13 writings, and finally agreed to count the sentences relevant to the questions 'Why it is necessary to consume wisely' and 'How will you be a smart consumer' as the key points (see Table 2 for examples). The answer to the first question may be a description of young people's current consumption style, an examination of the problems caused by irrational consumption, or a reflection of one's own experience of wise or unwise consumption. For the second question, we deemed the suggestions or specific steps to develop a reasonable consumption habit to be key points. We regarded the number of key points as the measure of content richness. In addition, the holistic score was given in accordance with the scoring criteria (see Appendix B) of the writing part of CET-4. Both the pilot coding of key points and the pilot scoring of the 13 essays were conducted, which reached a high agreement, as indicated by the Pearson coefficient ($r = .776$ for the pilot coding of key points; $r = .812$ for the pilot scoring). Then, the first author coded and scored the rest.

We analysed the participants' rating results of the 5-point scales on the questionnaires via independent samples t-tests to detect whether their perceptions of the effects of the prewriting discussion differed based on the language used. In addition, their written

Table 1
Indices of syntactic complexity.

	Full name	Description
Factor 1 phrasal level		
VPD	verb phrase density	the number of verb phrases per 1000 words
PPD	preposition phrase density	the number of preposition phrases per 1000 words
Factor 2 sentence level		
MSL	mean sentence length	the average number of words in each sentence
LE	left embeddedness	the mean number of words before the main verb of the main clause in sentences
Factor 3 noun phrase use		
MNP	modifiers per noun phrase	the mean number of modifiers per noun phrase
NPD	noun phrase density	the number of noun phrase per 1000 words

Table 2
Examples of the key points identified in the study.

Why it is necessary to consume wisely?
College life has become independent. Food, clothing and travel are arranged by ourselves, so I think it's important to have a bill to record my consumption. (2 key points counted)
Nowadays, many people don't have a reasonable budget and have to borrow money, put themselves in terrible debt eventually. (1 key point counted)
There are many students stuck in situation where they lack money just because of failing to budget their money. (1 key point counted)
… …
How will you be a smart consumer?
Usually, I'll divide the total money into monthly and daily money. Then, I'll record my daily consumption on a software so that I can clearly get the whereabouts of money from it.(2 key points counted)
As far as I'm concerned, what we need to do firstly is calculate how much we spend per *month so that we can find where we cost too much money. (1 key point counted)*
In the first place, make a list of daily necessities we need, clothing, and even food to make sure we spend within our means. (1 key point counted)
… …

feedback was coded. We created the coding framework by abstracting the texts into words and phrases, categorising these words and phrases according to the topics they represent, and then refining and developing a full list of categories that could cover all the topics entailed in the feedback without overlap or redundancy (Bogdan & Biklen, 2007).

To better interpret the findings of the effects of L1 and L2 prewriting discussions on writing performance, we also examined the discussions. Although each group had 10 min for their respective discussions, the actual average time of the discussions in the students' L1 (*Mean* = 5′03″; *SD* = 1′15″) was shorter than in their L2 (*Mean* = 8′24″; *SD* = 1′83″). We transcribed the recordings of the discussions using an online transcribing platform supported by IFLYTEK. By comparing each speaker's number of turns and his or her partner's in their pair discussion, we found no dominant speaker in either group, and the discussions displayed a relatively even pattern of language contribution. In addition, following the analytical framework of discussion focus, which was proposed by Storch (2005) and then adopted by Liao (2018), we examined three focus areas: (1) generating ideas, (2) lexicon-related episodes, and (3) grammar-related episodes. We found that in both groups, participants focused on generating ideas in their discussions, with a mean of 83.55% (*SD* = 4.45%) turns in their L1 discussion and 82.23% (*SD* = 5.33%) turns in their L2 discussion. Neither group spared much effort in negotiating words or grammar, and they did not mix their use of L1 and L2 throughout their discussions. For the L2 group, their EFL proficiency could support their communication, even though the words and sentence structures used in their discussions were simple. For the L1 group, one participant used her L2 when she answered the question asked by her partner: What is the English word for '*li xing xiao fei*' (rational consumption)?

4. Results

4.1. Results of textual analysis and holistic score

Table 3 summarises the writing performance of each group as indicated by CAF, content richness, and holistic score. In general, by comparing the means of different measures, we found that the writing quality of the L1 group was better than that of the L2 group across almost all measures, even though there was no significant difference in most measures. In particular, there was a significant difference with a medium effect size on three measures: (1) MTLD, which indicates lexical complexity, $t\,(63) = 2.08$, $p = 0.042$, $d = 0.52$; (2) accuracy suggested by the ratio of linguistic errors, $t\,(63) = -2.66$, $p = 0.01$, $d = -0.63$; and (3) fluency, denoted by the total number of words, $t\,(63) = 2.02$, $p = 0.047$, $d = 0.51$. In addition, the participants in the L1 group scored higher holistically than the L2

Table 3
Descriptive statistics of the writing quality of two groups.

	L1 group ($N = 35$)		L2 group ($N = 30$)	
	Mean	*SD*	*Mean*	*SD*
Complexity				
VPD	270.77	34.51	259.80	29.85
PPD	77.66	21.31	78.90	15.34
MSL	14.25	2.88	13.39	2.80
LE	3.93	1.57	3.85	1.37
MNP	0.68	0.16	0.62	0.10
NPD	337.98	33.21	341.04	34.60
MTLD	99.23	25.77	85.91	25.71
Accuracy	0.35	0.18	0.47	0.20
Fluency	208.49	38.73	189.80	35.13
Content richness	10.17	3.00	9.90	3.68
Holistic score	10.21	1.59	9.57	1.38

Note. VPD = verb phrase density; PPD = preposition phrase density; MSL = mean sentence length; LE = left embeddedness; MNP = modifiers per noun phrase; NPD = noun phrase density; MTLD = the measure of textual lexical density.

group, and although the p value did not suggest a conventionally significant level of difference, the medium effect size was still notable, $t\,(63) = 1.74$, $p = 0.087$, $d = 0.43$.

4.2. Results of participants' ratings and feedback

Table 4 illustrates the participants' rating results, which reflect their views of the usefulness of the prewriting discussion. We detected a trend approaching a significant difference with a medium effect size in their ratings on fluency improvement, $t\,(49) = 1.95$, $p = 0.057$, $d = 0.54$. This indicates that the participants who spoke in their L1 before writing attributed their writing fluency to the prewriting discussion to a larger extent; indeed, they wrote more words than the L2 group according to the outcomes of textual analysis. For other aspects, there was no significant difference.

The participants' feedback on their perceptions concerning the effect of the prewriting discussion was different. While we did not find any completely negative comments in the L1 group, four participants in the L2 group totally denied the role of the prewriting discussion. In general, the participants in the L1 group were more willing to report the advantages of the prewriting discussion, while the participants in the L2 group were more prone to reporting the disadvantages. Table 5 lists participants' views on the effect of the prewriting discussion; a more specific discussion of their feedback is presented in the next section.

5. Discussion

The results were interpreted primarily under the model proposed by Kellogg (1996) and Kellogg et al. (2013), which is commonly used in L2 writing studies and closely related to our study. Kellogg's model links writing processes to working memory systems, and offers a lens for us to analyse the cognitive activities of participants in different groups.

Within the framework of L1 writing production outlined by Kellogg, there are three basic processes that recursively occur along with the writing process: (1) formulation, (2) execution, and (3) monitoring. Kellogg related each of these processes to the resources of working memory outlined by Baddeley (1986), and indicated that the production of a text relies heavily on the capacity of working memory. Although this was initially used to interpret the L1 writing process, many L2 writing studies have applied the model and provided support for its main assumptions (e.g., Ellis & Yuan, 2004; Johnson et al., 2012). In L2 writing planning studies, the formulation process is of great interest. According to Kellogg, formulation incorporates two sub-processes: (1) planning, during which the writer generates and organises ideas related to writing goals; and (2) translating, which is a process of phonologically, orthographically, and grammatically encoding the ideas generated before. We believe that prewriting discussion may greatly benefit the formulation process since through discussion, writers can exchange individual ideas related to the topic and negotiate the use of words and grammar.

5.1. Research question 1: the effect on textual features and holistic score

Research Question 1 asks about the effects of L1 and L2 prewriting discussions on the participants' writings. The results of the textual analysis demonstrated that they did have different effects in that L1 discussions were more effective in improving writing quality in terms of lexical complexity, accuracy, fluency, and holistic score, with the exception of syntactic complexity and content richness.

One reason for the findings might be that compared to L2 discussions, L1 prewriting discussions played a more effective role in helping writers in their planning process, thus saving writers' time and attentional resources to attend to other aspects of writing, such as translating and monitoring. Prewriting discussions, as a form of collaborative dialogue, could permit writers to co-build the knowledge they need in subsequent writing. According to our analysis of the discussions, the participants in both groups devoted most of their time to generating ideas, which prepared them in the planning stage when composing their individual writings. However, the difference was that participants in the L1 group may have exchanged their ideas in a more efficient way with clearer information, less time spent, and less attentional resources consumed. As intermediate L2 learners, they were bound to feel more comfortable when talking in their native language and to rely more on their L1 to generate ideas (Wang & Wen, 2002). Compared with the participants in the L2 group, they saved more time and working memory capacity in their planning process; this may have improved their performance in other processes, such as translating and monitoring, which are directly linked with writing quality because translating concerns the linguistic encoding of conceptions, and monitoring is a reviewing process for correcting various errors.

Another reason could have been that during the prewriting discussion, the participants in the L2 group did not gain more

Table 4
Participants' ratings of the effect of prewriting discussions.

	L1 group ($N = 28$)		L2 group ($N = 23$)	
	Mean	*SD*	*Mean*	*SD*
Complexity	2.54	0.69	2.61	1.31
Accuracy	2.57	1.03	2.74	1.10
Fluency	3.64	0.91	3.09	1.13
Content richness	3.96	1.00	3.52	0.90
Holistic quality	3.25	0.80	3.04	1.11

Table 5
Feedbacks of the participants (translated from Chinese).

L1 group	L2 group
● It helped organize ideas and improve writing fluency (8/28) ● It enabled me to enrich writing content (21/28) ● I learnt new words through discussion (2/28) ● I lost my original ideas after discussion (1/28) ● It had a limited effect on improving linguistic complexity (3/28)	● It helped organize ideas and improve writing fluency (6/23) ● It enabled me to enrich writing content (15/23) ● Sentence structures, grammars, and words were discussed and learnt (2/23) ● Most ideas in the composition were out of my own mind (3/23) ● The discussion was not fruitful due to the limited oral English (2/23) ● It had a limited effect on improving linguistic complexity (1/23) ● It was not useful at all (4/23)

advantages over the L1 group for the translating process due to their limited L2 proficiency. It could be predicted based on Kellogg's model that L2 discussions more heavily favour the writers' linguistic encoding process since participants can rehearse most of the sentences in the discussion before writing them down in texts; further, they can gain new L2 knowledge through mutual learning and correction of L2 lexical and syntactic use. However, due to limited L2 oral proficiency, most participants could only produce simple, short utterances with errors. Thus, they still needed to spend time and attention encoding their ideas into grammatically complex and correct sentences to gain a high score. Even though some of them had negotiated lexical and grammatical use in L2, they were still unable to transfer the discussion outcome to their individual writings. For example, in Excerpt 1, student A corrected student B's use of 'going to shopping' by repeating the phrase in its correct form; however, student B still misused it in the post-discussion writing and did not pick up on the word 'proportion' used by student A. Moreover, this failure to transfer form-focused knowledge may have faded if the participants were more proficient in L2.

Excerpt 1.

	A/B discussion	B post-discussion writing
A:	*I like reading books, so I usually spend a large proportion buying books. How about you?*	Going to shopping takes a lot money from me …
B:	*Hmm. I think I like go to shopping, maybe.*	
A:	*Go shopping? you mean?*	
B:	*Yeah, and I always spend a lot. That's really bad.*	

Although the participants in the L1 group outperformed their peers in the L2 group in many aspects of writing quality, the measures of syntactic complexity and content richness did not display a significant difference. In other words, the difference in language used in the prewriting discussions did not lead to a different level of syntactic complexity and content richness of the post-discussion texts. The analysis of the discussions showed that both groups were focused on formulating ideas with almost no grammar-related episodes. Hence, their performance on syntactic complexity might be more attributable to their similar L2 writing proficiency, as indicated by their writing scores on the mid-term exam. Although the participants in the L1 group could use more working memory capacity during the translating process, they did not perform better in transforming their ideas into highly sophisticated sentences due to their limited knowledge of L2 syntax. Regarding the finding that there was no significant difference in content richness between the two groups' writings, it is possible that sufficient time bridged the gap between the role of the L1 and L2 discussions in enriching writing content. With a similar number of key points in their writings, the participants in the L2 group needed more discussion time to generate these ideas; hence, the current findings might change if the maximum discussion time for the L2 group were to be reduced. Another possible reason lies in the measure of content richness. Although the number of sentences answering the key questions predetermined was thought to be more accurate as a measure of content quality than the subjective rating, it might not have been effective enough in distinguishing the richness of ideas in different essays. As such, devising a more rigorous method of gauging the content richness of L2 writings is urgent in future studies.

Our central result stands in contrast to that of the previous study by Pu (2009). Pu found that writings after L2 prewriting discussions were superior in syntactic complexity, accuracy, and holistic score, even though no significant difference was displayed in these measures given the small sample size. The conflicting findings between the work of Pu and our study are probably attributable to the participants' academic backgrounds. Although both studies involved first-year university students, the participants in Pu's study were English majors with higher English proficiency than the participants in our study. It is possible that highly proficient L2 learners could produce more complex sentences with fewer errors in their L2 discussions; this supplies abundant resources regarding L2 vocabulary and grammar, thus resulting in better writing in terms of syntactic complexity and accuracy. In contrast, learners with lower L2 proficiency have a limited pool of resources in their L2 (Kellogg, 1996), so they might not benefit too much from each other in their L2 discussions.

Our findings also echo previous research, which has claimed that students' L1 might be a useful tool in form-focused tasks (e.g., Gánem-Gutiérrez & Roehr, 2011; Scott & de la Fuente, 2008). For example, in the study of Scott and de la Fuente (2008), both intermediate-level learners of Spanish and French performed better on an awareness-raising grammar task when they were allowed to use their L1 to complete the task. For this task, they were required to articulate a grammar rule after observing the texts that illustrated the targeted form. It may be that the discussion in the L1 was more balanced and coherent, and as such reduced the cognitive load for

this type of task. Although we did not adopt a form-focused task, the prewriting discussion in the students' L1 was also more efficient in that it better prepared the writers in the planning stage of individual writing, and saved them more cognitive resources for linguistic encoding and monitoring. The finding that the error ratio was lower in the writings of the L1 group indicated that L1 discussions might indirectly benefit writers' form-focused performance by saving cognitive efforts devoted to meaning-related problems. The findings also corroborate the claim that L1 use is not necessarily detrimental in L2 writing (e.g., DiCamilla & Antón, 2012; Wang & Wen, 2002; Wei et al., 2020; Zhang, 2018). However, as claimed by DiCamilla and Antón (2012), advanced L2 learners might prefer using their L2 as a mediating tool for communication; therefore, future studies are expected to involve highly proficient L2 learners when comparing the effects of L1 and L2 use on L2 writing.

5.2. *Research question 2: the participants' perceptions*

Research question 2 investigates the participants' perceptions of the effects of their prewriting discussion. Given their ratings, we proposed that the participants of the L1 group would be more positive about the effects of the prewriting discussion, especially in improving writing fluency. This was basically consistent with the textual analysis of participants' writings, which further proved the different effects of the prewriting discussions in L1 and L2.

Regarding the participants' written feedback, we found two differences in their perceptions between the L1 and L2 groups. First, three out of 23 participants in the L2 group reported that the discussion did not help with the content of their writings, in which most ideas were from their own minds; in contrast, we did not observe such comments in the feedback of the L1 group. Instead, one of 28 participants in the L1 group commented that he was excessively influenced by his partner in the discussion and eventually lost track of some of his original views. These distinctions in perceptions corroborate the hypothesis that the prewriting discussion in students' L1 is more helpful in preparing intermediate-level L2 writers in their planning process, because in using their native language, writers can internalise knowledge from their peers in a more successful way. However, the comment from the L1 group also raised a new question for future studies regarding whether prewriting discussion will impair writers' originality and creativity, and whether the results will be different when the discussion is conducted in a different language (i.e., L1 or L2).

Another difference was that four out of 23 participants in the L2 group admitted that the limited effect of their prewriting discussions was due to their limited spoken English proficiency; this was not true for the L1 group. To some extent, this difference in the participants' comments helped to explain the outcome whereby the writings after the discussions in L2 were not as good as the writings after the L1 discussions.

The perceptions of the L1 and L2 groups also have something in common. In both groups, most participants spoke highly of the effect on enriching the writing content, and they all performed well on this aspect accordingly. Based on sociocultural theory (Lantolf et al., 2015; Vygotsky, 1978), language can effectively mediate the process of learners internalising external knowledge. The participants' common perceptions suggest that even though L1 and L2 may have different mediating effects, they could still co-build knowledge in their collaborative dialogue (Swain, 2000) and benefit from it during their writing process. As most participants reported in their feedback, since they had different backgrounds and life experiences, they could obtain much inspiration and different perspectives from others, which was helpful in enriching the writing content.

Another issue is that among the four participants who negated the effects of the prewriting discussions in the L2 group, one participant mentioned that her partner was not active enough, and the interaction was not helpful in inspiring new ideas. A closer look into the participant's discussion with her partner revealed that although the two speakers were balanced in their turns, the other speaker tended to merely repeat the participant's ideas most of the time. Even though this was rare in this study, it indicated that discussion may be less collaborative when there is a huge gap between the two participants' knowledge, or when one of the speakers is reluctant to contribute. In this case, the instructor's intervention may be essential.

6. Conclusion

We investigated the effects of prewriting discussions in L1 and L2 on individual L2 writings and L2 learners' perceptions of the effects. The quantitative analysis suggests that the L1 prewriting discussion has some advantages in improving lexical complexity, accuracy and fluency. The results of the learners' ratings and feedback indicate that the participants who spoke in their L1 were generally more positive about the effects of their discussions. Although we only employed a one-time experiment, our findings imply the possibility that prewriting interaction with peers could facilitate the long-term development of learners' writing skills, which is worth further research.

Our study has significant implications for both future research and L2 writing pedagogy. First, we advanced from the research on the nature of L1 use in the L2 writing process to the investigation of the practical effects of L1 discussions prior to L2 writing. Pedagogically, we found that the use of L1 in L2 writing classrooms should be treated with caution, and at the very least should not be completely avoided. According to our findings, the L1 prewriting discussion was more productive, with fewer cognitive resources consumed in idea formulation, thus leading to greater writing fluency, better facilitating the linguistic encoding of the ideas generated, and benefitting the monitoring process for higher writing accuracy. However, our findings might not be true for some advanced L2 learners, who are able to manage their L2 discussions without making too much cognitive effort. Therefore, the instructor of L2 writing should provide balanced opportunities for both L1 and L2 prewriting discussions in a way that caters to the specific group of L2 learners. In addition, making learners aware of the potential role of L1 use in L2 writing, and cultivating their awareness to compare the different L1 and L2 rhetorical knowledge, may help prevent negative transfer (Wei, 2020). Second, some participants reported that the exchange of ideas might lead to the loss of their original views, which did not benefit their writing creativity. To further investigate

how much information in their discussions was transferred to their writings and how much was not, future studies could compare their discussion transcripts and writings in detail to identify the transfer patterns or adopt the post-writing retrospective questionnaire used in Wei et al. (2020) to probe learners' own perceptions of their L1-to-L2 transfer. Third, one participant in the L2 group who offered negative feedback complained that the partner was not active enough, which led to an unsuccessful discussion. There is a possibility that the disparity between the spoken English proficiency of participants in one pair was so large that they could not communicate effectively. Therefore, determining the appropriate difference in spoken English level between two learners of one discussion pair is of vital importance. This issue also posed a challenge for the language instructor, who is expected to monitor the discussion process, motivate active participation, and offer timely help. Finally, with the development of information and communication technology, online learning has played an indispensable role in school education, and providing explicit interactive activities to learners unfamiliar with each other online may promote their acquisition of sophisticated cognitive skills such as complex language skills (Jiang & Zhang, 2020). Therefore, it is also necessary to investigate whether prewriting discussion in online learning environments will produce different effects on individual writing performance.

Our study also has some limitations. First, we adopted a between-group design that minimised the practice effect across conditions; however, the individual differences among participants in different conditions might have influenced the outcome. Therefore, future studies should employ a within-group design to verify our findings by testing the same group of participants in all conditions (i.e., L1 prewriting discussion versus L2 prewriting discussion). Second, even though the instructor specified the language that participants should use in their discussion, there was no reminder that participants could negotiate grammatical and lexical issues apart from exchanging opinions on the topic. In this case, most participants might not pay enough attention to talking about L2 lexical and syntactic use. Future studies should continue to explore whether the results will change if the instructor permits the mixed use of L1 and L2 in prewriting discussion, and remind them to discuss the usage of sentences and words. In addition, the closed-ended questions in the questionnaire used in this study might be not effective and detailed enough to portray participants' perceptions. A questionnaire with more items targeting each sub-construct should be applied in future studies to ensure higher validity. Finally, the unequal proportion of male and female participants might influence the accuracy of the results since discussion is a social activity, and different combinations of genders in pairs could produce different effects on motivation and concentration.

Funding sources

This work was supported by the National Social Science Foundation of China [19CYY050].

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. The prompt of the independent writing task

Too much spending and borrowing money can easily put you in debt. Do you think it is necessary to be a smart consumer, and how will you be a smart consumer? Use reasons and examples to support your answers.

Appendix B. Questions used in the questionnaire

1 To what extent do you think the discussion improved the complexity of the sentence structures and vocabulary in your writing?
2 To what extent do you think the discussion improved the accuracy of grammar and vocabulary in your writing?
3 To what extent do you think the discussion improved your writing speed and fluency?
4 To what extent do you think the discussion improved the content richness of your writing?
5 To what extent do you think the discussion improved the holistic quality of your writing?
6 Please write down your feelings about the discussion. If you find it helpful, please write down in what ways. If you find the discussion unhelpful, or if it hinders your writing, please write about its shortcomings.

Appendix C. The scoring criteria of CET-4 writing section

The translation of the scoring criteria by the researchers.

Level (points)	Description
13–15	Relevant; contents well developed without ambiguity; grammar, spelling, and punctuation almost free from errors.
10–12	Relevant; contents well organized; a few errors in grammar, spelling, and punctuation, which do not interfere with comprehension.
7–9	Basically relevant; contents barely organized and not clear enough; frequent errors in grammar, spelling, and punctuation, some of which are serious errors that interfere with comprehension.
4–6	Basically relevant; contents not clear and badly organized; many serious errors in grammar, spelling and punctuation that interfere with comprehension.
1–3	Contents not clearly organized; fragmented sentences that abound with serious errors which interfere with comprehension.

References

Adams, R., Nik Mohd Alwi, N. A., & Newton, J. (2015). Task complexity effects on the complexity and accuracy of writing via text chat. *Journal of Second Language Writing, 29*, 64–81. https://doi.org/10.1016/j.jslw.2015.06.002

Al Masaeed, K. (2016). Judicious use of L1 in L2 Arabic speaking practice sessions. *Foreign Language Annals, 49*(4), 716–728. https://doi.org/10.1111/flan.12223

Althouse, A. D. (2016). Adjust for multiple comparisons? It's not that simple. *The Annals of Thoracic Surgery, 101*(5), 1644–1645. https://doi.org/10.1016/j.athoracsur.2015.11.024

Armstrong, K. M. (2010). Complexity, accuracy, and fluency in graded and ungraded writing. *Foreign Language Annals, 43*, 690–702. https://doi.org/10.1111/j.1944-9720.2010.01109.x

Aubrey, S., & Leeming, P. (2020). The impact of first as opposed to second language pre-task planning on the content of problem-solving task performance. *Language Teaching Research*, 1–26. https://doi.org/10.1177/1362168820917844

Azimzadeh, M. (2017). The effects of cognitively complex task on fluency, accuracy and complexity of L2 learners' written performance. *International Journal of Languages' Education and Teaching, 5*(3), 518–528. https://doi.org/10.18298/ijlet.1783

Baddeley, A. D. (1986). *Working memory*. Oxford University Press.

Bardovi-Harling, K., & Bofman, T. (1989). Attainment of syntactic and morphological accuracy by advanced language learners. *Studies in Second Language Acquisition, 11*, 17–34. https://doi.org/10.1017/S0272263100007816

Bogdan, R. C., & Biklen, S. K. (2007). *Qualitative research for education: An introduction to theories and methods* (5th ed.). Pearson.

Cheung, Y. L. (2016). A comparative study of paper-and-pen versus computer-delivered assessment modes on students' writing quality: A Singapore study. *The Asia-Pacific Education Researcher, 25*, 23–33. https://doi.org/10.1007/s40299-015-0229-2

Cho, M. (2018). Task complexity, modality, and working memory in L2 task performance. *System, 72*, 85–98. https://doi.org/10.1016/j.system.2017.10.010

Cohen, A. D., & Brooks-Carson, A. (2001). Research on direct versus translated writing: Students' strategies and their results. *The Modern Language Journal, 85*, 169–188. https://doi.org/10.1111/0026-7902.00103

Crossley, S. A., & McNamara, D. S. (2014). Does writing development equal writing quality? A computational investigation of syntactic complexity in L2 learners. *Journal of Second Language Writing, 26*, 66–79. https://doi.org/10.1016/j.jslw.2014.09.006

Cumming, A. (1990). Metalinguistic and ideational thinking in second language composing. *Written Communication, 7*, 482–511. https://doi.org/10.1177/0741088390007004003

DiCamilla, F. J., & Antón, M. (2012). Functions of L1 in the collaborative interaction of beginning and advanced second language learners. *International Journal of Applied Linguistics, 22*(2), 160–188. https://doi.org/10.1111/j.1473-4192.2011.00302.x

Ellis, R., & Yuan, F. Y. (2004). The effects of planning on fluency, complexity, and accuracy in second language narrative writing. *Studies in Second Language Acquisition, 26*(1), 59–84. https://doi.org/10.1017/s0272263104261034

Feise, R. J. (2002). Do multiple outcome measures require p-value adjustment? *BMC Medical Research Methodology, 2*, 1–4. https://doi.org/10.1186/1471-2288-2-8

Frear, M. W., & Bitchener, J. (2015). The effects of cognitive task complexity on writing complexity. *Journal of Second Language Writing, 30*(4), 45–57. https://doi.org/10.1016/j.jslw.2015.08.009

Frost, K., Elder, C., & Wigglesworth, G. (2011). Investigating the validity of an integrated listening-speaking task: A discourse-based analysis of test takers' oral performances. *Language Testing, 29*(3), 345–369. https://doi.org/10.1177/0265532211424479

Gánem-Gutiérrez, G. A., & Roehr, K. (2011). Use of L1, metalanguage, and discourse markers: L2 learners' regulation during individual task performance. *International Journal of Applied Linguistics, 21*(3), 297–318. https://doi.org/10.1111/j.1473-4192.2010.00274.x

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavioural Research Methods, Instruments, and Computers, 36*, 193–202. https://doi.org/10.3758/bf03195564

Guo, C. J., & Liu, F. (1997). A dynamic research into L1 influence on L2 writing. *Modern Foreign Languages, 78*, 30–38. https://en.cnki.com.cn/Article_en/CJFDTotal-XDWY704.002.htm.

Ho, M., & Savignon, S. (2007). Face-to-face and computer-mediated peer review in EFL writing. *CALICO Journal, 24*(2), 269–290. https://doi.org/10.1558/cj.v24i2.269-290

Huang, Y., & Zhang, L. J. (2020). Does a process-genre approach help improve students' argumentative writing in English as a foreign language? Findings from an intervention study. *Reading & Writing Quarterly, 36*, 339–364. https://doi.org/10.1080/10573569.2019.1649223

Jiang, D., & Zhang, L. J. (2020). Collaborating with 'familiar' strangers in mobile-assisted environments: The effect of socializing activities on learning EFL writing. *Computers & Education, 150*, 1–14. https://doi.org/10.1016/j.compedu.2020.103841

Johnson, M. D. (2017). Cognitive task complexity and L2 written syntactic complexity, accuracy, lexical complexity, and fluency: A research synthesis and meta-analysis. *Journal of Second Language Writing, 37*, 13–38. https://doi.org/10.1016/j.jslw.2017.06.001

Johnson, M., Mercado, L., & Acevedo, A. (2012). The effect of planning sub-processes on L2 writing fluency, grammatical complexity, and lexical complexity. *Journal of Second Language Writing, 21*, 587–599. https://doi.org/10.1016/j.jslw.2012.05.011

Kagan, S. (1995). Group grades miss the mark. *Educational Leadership, 52*(8), 68–71.

Kang, S., & Lee, J. (2019). Are two heads always better than one? The effects of collaborative planning on L2 writing in relation to task complexity. *Journal of Second Language Writing, 45*, 61–72. https://doi.org/10.1016/j.jslw.2019.08.001

Kellogg, R. T. (1996). A model of working memory on writing. In C. M. Levy, & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 57–71). Erlbaum.

Kellogg, R. P., Whiteford, A. E., Turner, C., Cahill, M., & Mertens, A. (2013). Working memory in written composition: A progress report. *Journal of Writing Research, 5*(2), 159–190. https://doi.org/10.17239/jowr-2013.05.02.1

Kessler, M., Polio, C., Xu, C., & Hao, X. (2020). The effects of oral discussion and text chat on L2 Chinese writing. *Foreign Language Annals, 53*, 666–685. https://doi.org/10.1111/flan.12491

Lantolf, J., Thorne, S. L., & Poehner, M. (2015). Sociocultural theory and second language development. In B. van Patten, & J. Williams (Eds.), *Theories in second language acquisition* (pp. 207–226). Routledge.

Lay, N. (1982). Composing processes of adult ESL learners: A case study. *Tesol Quarterly, 16*, 406–407.
Liao, J. (2018). The impact of face-to-face oral discussion and online text-chat on L2 Chinese writing. *Journal of Second Language Writing, 41*, 27–40. https://doi.org/10.1016/j.jslw.2018.06.005
Li, H. H., & Zhang, L. J. (2021). Effects of structured small-group student talk as collaborative prewriting discussions on Chinese university EFL students' individual writing: A quasi-experimental study. *PLoS One, 16*(5), 1–20. https://doi.org/10.1371/journal.pone.0251569. e0251569.
Li, H. H., Zhang, L. J., & Parr, J. M. (2020). Small-group student talk before individual writing in tertiary English writing classrooms in China: Nature and insights. *Frontiers in Psychology, 11*(570575), 1–16. https://doi.org/10.3389/fpsyg.2020.570565
Lockhart, C., & Ng, P. (1995). Analyzing talk in ESL peer response groups: Stances, functions and content. *Language and Learning, 45*(4), 605–655. https://doi.org/10.1111/j.1467-1770.1995.tb00456.x
Lu, X. (2017). Automated measurement of syntactic complexity in corpus-based L2 writing research and implications for writing assessment. *Language Testing, 34*(4), 493–511. https://doi.org/10.1177/0265532217710675
Manchon, R. M., Roca, J., & Murphy, L. (2000). An approximation to the study of backtracking in L2 writing. *Learning and Instruction, 10*, 13–35. https://doi.org/10.1016/s0959-4752(99)00016-x
McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods, 42*, 381–392. https://doi.org/10.3758/BRM.42.2.381
McDonough, K., & De Vleeschauwer, J. (2019). Comparing the effect of collaborative and individual prewriting on EFL learners' writing development. *Journal of Second Language Writing, 44*, 123–130. https://doi.org/10.1016/j.jslw.2019.04.003
McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press. https://doi.org/10.1017/CBO9780511894664
McNamara, D. S., Louwerse, M. M., McCarthy, P. M., & Graesser, A. C. (2010). Coh-Metrix: Capturing linguistic features of cohesion. *Discourse Processes, 47*, 292–330. https://doi.org/10.1080/01638530902959943
Michel, M., Kormos, J., Brunfaut, T., & Ratajczak, M. (2019). The role of working memory in young second language learners' written performances. *Journal of Second Language Writing, 45*, 31–45. https://doi.org/10.1016/j.jslw.2019.03.002
Muhammad, R. (2019). Effects of increasing the degree of reasoning and the number of elements on L2 argumentative writing. *Language Teaching Research, 23*(5), 633–654. https://doi.org/10.1177/1362168818761465
Nelson, G. L., & Carson, J. G. (1998). ESL students' perceptions of effectiveness in peer response groups. *Journal of Second Language Writing, 7*(2), 113–131. https://doi.org/10.1016/S1060-3743(98)90010-8
Neumann, H. (2014). Teacher assessment of grammatical ability in second language academic writing: A case study. *Journal of Second Language Writing, 24*, 83–107. https://doi.org/10.1016/j.jslw.2014.04.002
Neumann, H., & McDonough, K. (2015). Exploring student interaction during collaborative prewriting discussions and its relationship to L2 writing. *Journal of Second Language Writing, 27*, 84–104. https://doi.org/10.1016/j.jslw.2014.09.009
Pae, T. (2018). Effects of task type and L2 proficiency on the relationship between L1 and L2 in reading and writing: An SEM approach. *Studies in Second Language Acquisition, 40*(1), 63–90. https://doi.org/10.1017/S0272263116000462
Plakans, L., Gebril, A., & Bilki, Z. (2016). Shaping a score: Complexity, accuracy, and fluency in integrated writing performances. *Language Testing, 36*(2), 1–19. https://doi.org/10.1177/0265532216669537
Pu, X. (2009). Effects of prewriting discussion on the language quality of argumentative writings. *Teaching English in China – CELEA Journal (Bimonthly), 32*(5), 16–24. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.627.2203&rep=rep1&type=pdf.
Rahimi, M., & Zhang, L. J. (2019). Writing task complexity, students' motivational beliefs, anxiety and their writing production in English as a second language. *Reading and Writing, 32*(3), 761–786. https://doi.org/10.1007/s11145-018-9887-9
Révész, A., Kourtali, N., & Mazgutova, D. (2017). Effects of task complexity on L2 writing behaviors and linguistic complexity: Task complexity and L2 writing. *Language Learning, 67*(1), 208–241. https://doi.org/10.1111/lang.12205
Roca, J., Murphy, L., & Manchon, R. (1999). The use of restructuring strategies in EFL writing: A study of Spanish learners of English as a foreign language. *Journal of Second Language Writing, 8*, 13–44. https://doi.org/10.1016/s1060-3743(99)80111-8
Ruiz-Funes, M. (2015). Exploring the potential of second/foreign language writing for language learning: The effects of task factors and learner variables. *Journal of Second Language Writing, 28*, 1–19. https://doi.org/10.1016/j.jslw.2015.02.001
Scott, V. M., & de la Fuente, M. J. (2008). What's the problem? L2 learners' use of the L1 during consciousness-raising form-focused tasks. *The Modern Language Journal, 92*(1), 100–113. https://doi.org/10.1111/j.1540-4781.2008.00689.x
Shi, L. (1998). Effects of prewriting discussions on adult ESL students' compositions. *Journal of Second Language Writing, 7*(3), 319–345. https://doi.org/10.1016/S1060-3743(98)90020-0
Storch, N. (2005). Collaborative writing: Product, process, and students' reflections. *Journal of Second Language Writing, 14*, 153–173. https://doi.org/10.1016/j.jslw.2005.05.002
Storch, N., & Aldosari, A. (2010). Learners' use of first language (Arabic) in pair work in an EFL class. *Language Teaching Research, 14*(4), 355–375. https://doi.org/10.1177/1362168810375362
Storch, N., & Wigglesworth, G. (2003). Is there a role for the use of the L1 in an L2 setting? *Tesol Quarterly, 37*, 760–769. https://doi.org/10.2307/3588224
Strauss, P., & U, A. (2007). Group assessments: Dilemmas facing lecturers in multicultural tertiary classrooms. *Higher Education Research and Development, 26*(2), 147–161. https://doi.org/10.1080/07294360701310789
Streiner, D. L. (2015). Best (but oft-forgotten) practices: The multiple problems of multiplicity-whether and how to correct for many statistical tests. *American Journal of Clinical Nutrition, 102*(4), 721–728. https://doi.org/10.3945/ajcn.115.113548
Swain, M. (1997). Collaborative dialogue: Its contribution to second language learning. *Revista Canaria de Estudios Ingleses, 34*, 115–132.
Swain, M. (2000). The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In J. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 97–114). Oxford University Press.
Swain, M., & Lapkin, S. (2000). Task-based second language learning: The uses of the first language. *Language Teaching Research, 4*, 251–274. https://doi.org/10.1177/136216880000400304
Thoms, J., Liao, J., & Szustak, A. (2005). The use of L1 in an L2 on-line chat activity. *The Canadian Modern Language Review, 62*(1), 161–182. https://doi.org/10.3138/cmlr.62.1.161
Tsagari, D., & Giannikas, C. N. (2018). Re-evaluating the use of the L1 in the L2 classroom: Students vs. teachers. *Applied Linguistics Review, 11*(1), 151–181. https://doi.org/10.1515/applirev-2017-0104
Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
Wang, W., & Wen, Q. (2002). L1 use in the L2 composing process: An exploratory study of 16 Chinese EFL writers. *Journal of Second Language Writing, 11*, 225–246. https://doi.org/10.1016/S1060-3743(02)00084-X
Wei, X. (2020). Assessing the metacognitive awareness relevant to L1-to-L2 rhetorical transfer in L2 writing: The cases of Chinese EFL writers across proficiency levels. *Assessing Writing, 44*(100452), 1–10. https://doi.org/10.1016/j.asw.2020.100452
van Weijen, D., van den Bergh, H., Rijlaarsdam, G., & Sanders, T. (2009). L1 use during L2 writing: An empirical study of a complex phenomenon. *Journal of Second Language Writing, 18*, 235–250. https://doi.org/10.1016/j.jslw.2009.06.003
Wei, X., Zhang, L. J., & Zhang, W. (2020). Associations of L1-to-L2 rhetorical transfer with L2 writers' perception of L2 writing difficulty and L2 writing proficiency. *Journal of English for Academic Purposes*, 1–14. https://doi.org/10.1016/j.jeap.2020.100907, 100907.
Wigglesworth, G., & Storch, N. (2009). Pair versus individual writing: Effects on fluency, complexity and accuracy. *Language Testing, 26*(3), 445–466. https://doi.org/10.1016/S1060-3743(02)00084-X

Woodall, B. (2002). Language-switching: Using the first language while writing in a second language. *Journal of Second Language Writing, 11*, 7–28. https://doi.org/10.1016/S1060-3743(01)00051-0

Zaccaron, R., & Xhafaj, D. C. P. (2020). Knowing me, knowing you: A comparative study on the effects of anonymous and conference peer feedback on the writing of learners of English as an additional language. *System, 95*, Article 102367. https://doi.org/10.1016/j.system.2020.102367

Zhang, M. (2018). Collaborative writing in the EFL classroom: The effects of L1 and L2 use. *System, 76*, 1–12. https://doi.org/10.1016/j.system.2018.04.009