

## Article

# The Role of Teacher-Generated, Learner-Generated, and Creative Content in Chinese EFL Students' Narrative Writing: A Contextual Perspective

Qiaoxia Wu <sup>1,\*</sup> and Agnes Albert <sup>2</sup> <sup>1</sup> School of Foreign Languages and Cultures, Ningxia University, Yinchuan 750021, China<sup>2</sup> Department of English Applied Linguistics, Eötvös Loránd University, 1088 Budapest, Hungary

\* Correspondence: wuqiaoxia0108@foxmail.com

**Abstract:** Task complexity has long been posited as an influential task feature inspiring much research. However, task complexity frameworks might be in need of adjustment, as they tend to emphasize the role of cognitive factors and neglect affective ones despite the fact that learner agency and potential for creativity have been linked to certain aspects of task performance, possibly exerting their influence through learners' affects. Thus, to investigate the role of agency and creativity in task-based L2 writing, this study aimed to explore the relationship between task conditions conceptualized as the levels of learner agency and the potential for creativity in Chinese students' English written performances on the one hand, and the possible role that the study contexts might play in the written performances in each task condition on the other. Participants of the study were two groups of Chinese intermediate learners of English studying in Hungary and China ( $n = 40$ ), producing 120 narratives altogether. In our study, different aspects of task performance, i.e., syntactic and lexical complexity and accuracy, were associated with learner agency and potential for creativity. Moreover, differences were found in fluency between Chinese students studying in the different contexts, indicating the possible role of study contexts in this regard.

**Keywords:** task complexity; task condition; learner agency; creative writing; written performance; fluency; lexical complexity; syntactic complexity; accuracy; study contexts



**Citation:** Wu, Qiaoxia, and Agnes Albert. 2022. The Role of Teacher-Generated, Learner-Generated, and Creative Content in Chinese EFL Students' Narrative Writing: A Contextual Perspective. *Languages* 7: 212. <https://doi.org/10.3390/languages7030212>

Academic Editor: Zsuzsanna I. Abrams

Received: 28 June 2022

Accepted: 4 August 2022

Published: 10 August 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Writing plays a major role in Second Language Acquisition (SLA) and has been suggested to be effective in promoting SLA (Kafipour et al. 2018; Polio 2012; Williams 2012). However, studies on writing instruction in high schools in China have shown that the current situation of English writing instruction is not ideal for the students. They highlighted several possible problems, one of which is that students are not interested in writing and find the writing classes boring (e.g., Chen 2015; Yu 2017; Zhang 2016). These circumstances call for a fine-grained investigation of the situation with the help of diverse tasks. For example, Maehr (1984) suggested that the socio-cultural context and prior personal experiences are important factors that determine the meaningfulness of educational tasks and situations for learners, which, in turn, determine learners' willingness to invest their personal resources, i.e., time, talent and energy, into completing a task. We argue that investigating writing with the help of different tasks could provide information about students' actual writing performance and might shed light on potential problems with writing tasks. Moreover, learning about the effects of certain task features, like exercising learner agency or the opportunity to be creative, might be beneficial for improving writing instruction in China.

There have been recent task-based studies linking learner agency and creativity to aspects of task performance: Fluency, complexity, and accuracy (e.g., Lambert et al. 2017; Phung 2017; Poupore 2014; Qiu and Lo 2017; Tin 2011, 2012). However, since studies

on these two task dimensions have mainly been conducted on oral tasks, it is doubtful whether their findings can be directly adopted to writing tasks. In light of this concern with transferability, writing instruction would benefit from information about how these task features impact learners' performance, making their research imperative.

Furthermore, many comparative studies have shown that the learning context also influences language learning outcomes (DeKeyser 2007; Segalowitz et al. 2004). However, research on study-abroad contexts was generally carried out in countries where the target language is spoken by the majority, that is, second language contexts. Few empirical studies have focused on contexts where English is not an official language but a lingua franca: The typical case of immigrants to Europe, for example. Therefore, the current study attempted to examine the role of exercising learner agency and potential for creativity in the English writing task performances of two groups of Chinese native speakers: One studying English in an English as a Lingua Franca (ELF) context in a non-English speaking country, Hungary, and the other studying English as a foreign language (EFL) in China.

## 2. Background

### 2.1. Chinese Students Learning English in China and Abroad

For Chinese students in China, English is one of the most important subjects since it accounts for 20% in the college entrance exam in most provinces. English writing is of even more importance in the sense that: (1) It takes up about 17% to 27% of the total score, varying in different provinces (see Yang 2020 for past exam papers in the recent college entrance exams), and (2) in China, due to geopolitical reasons, English learners have few opportunities to use English by speaking it. This is especially true in the western part of the Chinese mainland, which leaves writing as the only way of using English for these students.

However, exam-oriented English classes focus more on reading and grammar, while writing receives little attention (Chen 2018), and students get bored of the traditional *students write, teachers review* pattern, and they seldom practice writing outside of school (Chen 2018; Yuan 2012). Besides, in formal English writing instruction, the genre of argumentation and mostly teacher-controlled writing are practiced in high schools in China, while other forms of writing are rarely used. For example, learners are unlikely to write narratives about their own experiences (Chen 2018; Zhang 2012; Zhang 2016). Moreover, the training of writing is often scheduled intensively before the big exam, with a focus on imitating writing models and writing strategies, which probably does not contribute to the improvement of the students' writing ability greatly (Liu 2020).

ELF contexts are quite different from EFL contexts (Cogo and Jenkins 2010). Thus, the situation of Chinese students studying in Hungary (ELF context) is quite different from those studying in China (EFL context) in terms of their English learning experiences. For example, in the case of Chinese immigrant students studying English in Hungary, English can be used to communicate with people who speak different mother tongues in social settings, and English might also be used as the medium of instruction in academic settings (Cogo and Jenkins 2010; Seidlhofer 2005). Compared to students living in China, immigrants in Hungary have more opportunities to be exposed to an English-speaking environment and more opportunities for receiving English input and producing output, such as talking with foreign friends or classmates from other countries or going for summer camps in neighboring European countries. Additionally, schooling options are varied in Hungary (Öveges and Csizér 2018), with both state-financed bilingual and self-financed international schools being present, resulting in a greater variety in students' writing experiences due to encountering teachers from different cultural and educational backgrounds (for example, see Bennett 2019). Therefore, we expected differences between the writing performances of Chinese students studying in Hungary (CSH) and China (CSC).

## 2.2. Using Tasks to Investigate Writing Task Performance

According to an oft-cited definition, a task is “an activity in which meaning is primary; there is some relationship to the real world, task completion has some priority, and the assessment of task performance is in terms of task outcome” (Skehan 1998, p. 38). When considered from a psycholinguistic perspective, tasks, especially certain features of tasks, the most frequently researched ones being task complexity or difficulty, are believed to engage learners in various types of information processing that are useful for language acquisition (Ellis 2018). In a more general sense, it is assumed that certain task features, like those contributing to task complexity, can be linked to different aspects of task performance, which necessitates close scrutiny of relevant task features and performance measures as these provide invaluable insights for task design.

In line with this argument, a significant proportion of task-based L2 writing studies has focused on the cognitive dimensions of task design, specifically features contributing to the cognitive complexity of tasks and foreign language performance on them (e.g., Abrams 2019; Kormos and Trebits 2012; Kuiken and Vedder 2008; Ruiz-Funes 2015; Wang and Jin 2022). The most popular theories on task complexity are Robinson’s (2001a, 2001b, 2005) Triadic Componential Framework and Cognition Hypothesis and Skehan’s (2001) limited-attentional resources model. The main difference between these models lies in their understanding of the relationship between task complexity and learners’ linguistic performance. While Robinson claims that increased task complexity can be matched with greater accuracy and complexity by directing learners’ attention to their linguistic system and certain linguistic codes to meet the cognitively more complex conceptual demands if cognitive complexity was increased via a *resource-directing* dimension (e.g., reasoning demands), Skehan holds that a *trade-off effect* will occur between complexity, accuracy, and fluency, suggesting that certain task features make students focus on different performance areas. Interestingly, the task feature of *familiarity*, which was addressed in quite a few, not even necessarily task-based writing studies recently (Bui and Luo 2021; He and Shi 2012; Kessler et al. 2022; Yang and Kim 2020; Yoon 2017), had already been present in both Skehan’s (1998, 2001) and Robinson’s (2001a, 2001b, 2005) respective task sequencing frameworks. The fact task familiarity can be interpreted in different ways, for example, as familiarity with the task type and as familiarity with or having previous knowledge about the task content, might be responsible for the relatively novel interest in topic familiarity. It seems that familiarity with the task type was in the focus of attention earlier; this issue was investigated in a number of task repetition studies, which indicated that this type of familiarity was quite limited in its potential to explain learners’ performance on similar types of tasks unless repetition involved the whole task, that is both task type and content were familiar (Samuda and Bygate 2008). Recent topic or content familiarity studies appear to be much more successful in supporting the positive effects of task familiarity, mostly with regard to increasing the lexical complexity (Bui and Luo 2021; He and Shi 2012; Kessler et al. 2022; Yang and Kim 2020), in some cases the syntactic complexity (Kessler et al. 2022; Yoon 2017) and the fluency (Bui and Luo 2021; He and Shi 2012), and sometimes the accuracy of task performance (He and Shi 2012).

Since Robinson’s (2001b, 2005) and Skehan’s (2001) models are grounded in and mostly applied to studies of oral production, these frameworks might be limited in their ability to model the effects of task complexity in writing since task complexity might affect task performance differently in different modes, such as oral versus written (Tavakoli 2014). When it comes to writing, task complexity is supposed to be connected to the writing process. Although Kellogg’s (1996) model of writing is an attempt to account for the L1 writing process, based on the analogous use of Levelt’s (1993) model in L2 speech production (Kormos 2014), it can be expected that the processes proposed in the model might also apply to L2 writers, with perhaps more limited automaticity, gaps in competence, and interference from the L1. Kellogg’s model includes three phases of the writing process, namely formulation, execution, and monitoring. Formulation entails planning when the writer establishes goals for the writing and generates ideas to reach

these goals. It also includes organizing the structure of what the writer wishes to write, as well as selecting lexical and syntactic frames that are needed to encode those generated ideas to make the linguistic units ready for execution. During the next step, execution, the writer's ideas or plans are converted into production schemata for handwriting or typing, a process of executing or the actual production of sentences. The final stage, monitoring, consists of reading, revisions, and editing. Therefore, for foreign language writing tasks, task complexity should be understood and analyzed regarding the different stages of the composing process. To be more specific, since writing tasks set in motion a series of complex mental processes that a writer goes through during each step of composing, the task is believed to affect the allocation of students' limited attentional and memory resources, which is likely to influence the writing process at each of the three stages (Kormos 2012), thus leading to written texts with different qualities. However, research about writing along this line, no matter whether in L1 or L2, is rather scarce.

### 2.3. Learner Agency and Potential for Creativity in Tasks and Task Performance

Besides task complexity, studies have shown that other task dimensions also seem to be linked to task engagement and task performance. For instance, studies have found that learner agency is likely to be linked to task engagement and different aspects of task performance.

According to Vygotsky (1978), learner agency consists of learners' sense of agency in general or concerning the particular context, as well as their agentic behaviors in the sense of acting or non-acting. Only when learners actively perceive the resources and contexts provided and interact with them does agency emerge, or can they be regarded as agentic learners (van Lier 2004, 2008). In the applied linguistics field, agentic language learners are defined as intentional learners who "assume responsibility for managing their own learning: Setting targets, making choices, making decisions, monitoring progress, and evaluating outcomes" (Little and Legenhausen 2017, p. 43). Researchers adopt different definitions of learner agency based on their own research goals, either in relation to motivation and learning strategies (Jin and Wang 2021) or students' selection of language learning and teaching materials (Matsumoto 2021). What should be noted is that in most recent task-based foreign language studies, learner agency seems to be conceptualized as how much control learners have over task content (Lambert and Minn 2007; Lambert et al. 2017). Studies have found that learners are likely to be more engaged and invested in tasks operated under the learner-generated content condition, where learners can create their own ideas and contents to express (for example, asking students to think of an interesting or funny story about a problem they personally experienced in the past that their partners would enjoy listening to, a task used in Lambert et al. 2017), than the teacher-generated content condition, where the content is predetermined by the teacher or the task designer (such as picture description, e.g., Lambert and Minn 2007; Lambert et al. 2017; Phung 2017; Poupore 2014; Qiu and Lo 2017), which calls for personally-invested L2 task design. Nevertheless, except for Lo and Hyland (2007), few empirical studies have been conducted on its relevance to writing. The present study is expected to broaden the scope of EFL learning research by investigating the relationship between the task dimension of learner agency and writing.

In addition to learner agency, the potential for creativity is another task feature that has been found to influence students' L2 performance (Albert and Kormos 2004; Kim et al. 2021; Tin 2011). In applied linguistics, creativity can be defined "as the playful use of language to construct new/unknown meanings, transforming one's current linguistic and conceptual world and involving several types of creative thinking" (Tin 2011, p. 216). Tin (2012) argues that creative tasks with constraints that require learners to construct new meanings or "unknown meaning" (p. 178), that is "constraint desirable for creativity" (p. 179), can facilitate creative language use resulting in the production of more complex language. Constraint is a key feature of a creative task, referring to "any restriction on freedom that limits the number of possible solutions available for solving the problem at

hand, including rules, goals, and limitations on choice, boundaries, and scarcity" (Joyce 2009, p. 5). In the context of classroom EFL, writing tasks of this kind are most likely to enhance language task performance as well as writing abilities (Abrams 2019; Maley 2006; Tin 2012).

In the current study, two task dimensions, learner agency and the potential for creativity, were investigated by analyzing performance on three types of tasks: (1) Content was provided by the teacher (thus it was predetermined for the students with little room for creativity), (2) students had the opportunity to use their own personal experiences and control the task content (therefore they could also use their creativity if they chose to, but the task intended to draw on their personal experiences rather than on their creativity), and (3) the task entailed some freedom in content, but with specific constraints, urging learners to come up with creative solutions. Although (2) and (3) appear quite similar in terms of allowing students to be creative since there is no predetermined content to check learners' solutions against, according to Tin (2011, 2012), the complete freedom and lack of constraints in (2) might also result in a mundane account of past experiences, possibly recounted numerous times earlier by the students. Therefore, creativity is only an option but certainly not a requirement there since accounts of everyday personal experiences can be quite ordinary.

### 3. Methods

#### 3.1. Research Questions

Based on the research background and research goals stated above, the following research questions were posed:

1. How do different task conditions (teacher-generated, learner-generated, and creative writing) relate to students' written performance in terms of the amount of output, lexical and syntactic complexity, and accuracy for CSH?
2. How do different task conditions (teacher-generated, learner-generated, and creative writing) relate to students' written performance in terms of the amount of output, lexical and syntactic complexity, and accuracy for CSC?
3. What possible role do the English learning contexts play in students' written performances in each task condition? That is, what kinds of differences might be detected between CSH and CSC in their written performances?

#### 3.2. Participants

Two groups of CSH (10 males and 10 females) and CSC (8 males and 12 females) participated in the study and were recruited through convenience sampling. Their ages ranged from 12 to 17 years of age ( $\text{Mean}_{\text{CSH}} = 13.75$ ,  $\text{SD}_{\text{CSH}} = 1.895$ ,  $\text{Mean}_{\text{CSC}} = 15.95$ ,  $\text{SD}_{\text{CSC}} = 0.06$ ). The difference in means is a result of difficulties in recruiting Chinese teenagers in Hungary. Nevertheless, in order to ensure that the samples were comparable, participants were selected based on their results (from 120 to 149 out of 200 points in total) on the Oxford Placement Test 1, New 2004 edition (OPT) in a way that their proficiency levels ranged from B1 to B2 according to the Common European Framework Reference (CEFR). The OPT test consists of two parts, *Listening* and *Use of English* (each accounting for 100 points), with the *Listening* part assessing listening skills while the *Use of English* focuses on grammar. It is also important to mention that CSH were expected to have been studying in Hungary for at least 3 years by the time of data collection so the aim of examining the possible role of the Hungarian English learning context on their English writing could be achieved.

Although members of the two groups were selected on the basis of their total proficiency scores obtained at least one week before administrating the tasks, it was revealed that the resulting two groups of Chinese students were significantly different in their total placement test scores ( $\text{Mean}_{\text{CSH}} = 140.10$ ,  $\text{Mean}_{\text{CSC}} = 130.45$ ,  $p < 0.05$ ). However, further *t*-tests showed that this significant difference was caused by the much higher *Listening* scores of students studying in Hungary, while the *Use of English* scores of the two groups ( $\text{Mean}_{\text{CHS}} = 61.15$ ,  $\text{Mean}_{\text{CSC}} = 64.50$ ,  $p > 0.05$ ) were not significantly different. Since the



present study was conducted on writing, where grammar reflected by the *Use of English* scores probably had a greater impact while listening skills did less so or not at all (Kormos and Trebits 2012), it is safe to say that the two groups of Chinese students were comparable regarding the overall effect of their English learning contexts on their written performance.

### 3.3. Instruments

Three written narrative tasks were used for data collection: A picture narration task, a personal experience narration task, and a story-creating task (see Appendix A for the three tasks), which were all designed by the authors. The three tasks were piloted with three students studying in Hungary who had intermediate English language proficiency. During the piloting, we attempted to ensure that the tasks, in fact, elicited narratives, and details such as the time needed for task completion and minimum length requirements were checked as well. Due to minor modification to the tasks as a result of piloting, the three students were excluded from the final study.

The picture narration task prompted the students to tell a story based on six related pictures in order. Since the content was predetermined by the teacher, this task was considered to be a teacher-generated content task (TGT). The personal experience narration task required the students to recall and write about a memorable event that happened in their lives. The students were in absolute control and could generate the task content themselves. Therefore, it cannot be ruled out that they might have invented their stories in response to this prompt (as it is the case whenever learners are asked to give a personal account of something). Our intention was nevertheless to give control to the students over task content and enable them to use their personal experiences, so it was labeled as a learner-generated content task (LGT). The creative task (CT) introduced some constraints while also providing some control for the students. Ten unrelated words were given to them, and the students were required to create a story by using at least six of them. The three tasks were different in the level of learner agency and the potential for creative language use. The TGT had the lowest level of learner agency, and the LGT had the highest, whereas the CT had the greatest potential to elicit creativity, according to the arguments proposed by Tin (2011, 2012), while the TGT had very limited potential. The three tasks had the same title “A memorable event”, and the starting sentence was also offered by the prompt “This is something that I will never forget . . .”. The task instructions were in Chinese to eliminate a potential confounding variable (i.e., not understanding the task).

### 3.4. Procedures

Data collection in Hungary was carried out in a classroom environment by one of the researchers in a weekend private school, while data collection in China took place within the academic period of high schools in China in a school environment. The students were given 30 min to complete each task, and although the instructions specified the minimum word limit (100 words), no maximum length was given. All the data were collected by one of the researchers. For both groups of Chinese students, the three tasks were given to the participants at least one week apart, and the tasks for all the participants were given in different orders to minimize the effect of task order on students’ task performance. All essays were hand-written without access to dictionaries or digital tools. Altogether, 120 narratives were collected from the 40 participants (20 in each group), with each student performing three tasks and producing 3 narratives, which resulted in 60 narratives produced by each group.

### 3.5. Analyses

In the current study, students’ performance on the tasks was examined in line with the approach generally adopted in task-based studies: The performance areas of complexity (both syntactic and lexical), accuracy, and fluency (CAF) (Michel 2017) were targeted. When analyzing fluency in writing, in parallel with fluency in speech, it is usually interpreted as the length of text produced within a time limit (Larsen-Freeman 1978; Ong and Zhang 2010;

Wolfe-Quintero et al. 1998). Although Abdel Latif (2013) raised concerns about whether such a measure is in line with the traditional interpretation of fluency as the learner's ability to mobilize linguistic resources for real-time communication, Johnson (2017) found that when composing time is held constant, total words produced within a set time can serve as a strong metric for L2 writing fluency. Therefore, we measured writing fluency as the total number of words produced (TW) within the 30 min the learners had to complete the tasks.

Complexity was evaluated regarding two aspects, syntax and lexis. With regard to syntactic complexity, the T-unit was chosen as the basic unit of analysis, which is defined as a single independent clause plus any subordinate clauses attached to it or embedded in it (Norris and Ortega 2009). In an attempt to provide a comprehensive picture of the different aspects of syntactic complexity, the mean length of T-units (MLT) was used as a general measure of complexity, the mean number of clauses per T-unit (MNCT) was applied as a measure of subordination, while the mean length of clause (MLC) was utilized as a measure of sub-clausal complexity in this study, in accordance with the recommendations of Norris and Ortega (2009). The guidelines for identifying T-units and clauses were established based on Polio's (1997), Polio and Shea's (2014), and Lee's (2009) to make them as inclusive as possible.

As regards lexis, both text-internal and text-external lexical complexity were analyzed in our study (Skehan 2009). Text internal lexical complexity, reflecting how much variation can be found within the text, was measured by a version of the type-token ratio, the D-value, which takes text length into consideration when calculating this measure of lexical diversity (Meara and Miralpeix 2016). Text external lexical complexity, reflecting the ratio of difficult or rare words in the text, was measured by P\_Lex, which is a program that divides the text into 10-word segments and relies on the number of difficult or unusual words there for calculating lexical sophistication (Meara and Miralpeix 2016). Both variables can be calculated with the help of online software available at <http://www.lognostics.co.uk/> (accessed on 9 February 2019). Before entering the data for analysis in the software, the original handwritten texts were edited by correcting the wrong spellings (e.g., thier/their) and changing the contracted forms (e.g., I'm/I am) and abbreviations into their full forms (e.g., WWI/World War I), since, if not, P\_Lex software considers them as infrequent even though they are not when typed out in their right and full forms in the software. P-Lex was designed to analyze texts produced by learners and thus works best with texts that are not longer than 300 words. The minimum number of words needed to run D\_Tools is 50 words (Meara and Miralpeix 2016).

Regarding accuracy in L2 performance, researchers can opt for local or global measures. Local measures aim to track the use of designated grammatical features in the L2, such as verb and noun morphology, while global measures focus on the overall level of accuracy in an oral or written product. Local measures can be problematic considering that students do not acquire grammatical features in a concurrent way, and it is also difficult to ensure that a specific grammatical feature occurs frequently enough in the data to make its measurement meaningful. By contrast, global measures avoid these problems since they essentially divide the data into segments and calculate the error rate and are usually preferred by researchers (Wigglesworth and Foster 2008; Wolfe-Quintero et al. 1998). Therefore, global accuracy measures were selected in the present study, specifically: The ratio of error-free clauses to the total clauses (%EFC) and the ratio of error-free T-units to the total T-units (%EFT).

Wigglesworth and Foster (2008) proposed a further method for measuring accuracy, the weighted clause ratio (WCR). Its strength lies in the fact that it goes beyond the mere dichotomy of correct versus incorrect and combines all the errors in a global score by categorizing each clause according to the gravity of errors within it. Clauses fall into the following four categories: Entirely accurate clauses, level 1 clauses, level 2 clauses, and level 3 clauses. Entirely accurate clauses are completely error-free; level 1 clauses refer to those that have only minor errors (e.g., in morphosyntax) not compromising meaning; level 2 clauses are defined as those that contain serious errors (e.g., verb tense, word choice,

or word order), but the meaning is recoverable though not always obvious; and level 3 clauses constitute those that have very serious errors making the intended meaning far from obvious and only partly recoverable. These different level clauses are assigned scores 1, 0.8, 0.5, and 0.1, respectively. Afterwards, the scores of entirely accurate, level 1, level 2, and level 3 clauses are calculated by multiplying the number of clauses in each category by the specific score attributed to that level (1, 0.8, 0.5, and 0.1, respectively). Then, WCR can be calculated by dividing the sum of the clause scores by the total number of the clauses (see Foster and Wigglesworth 2016). WCR is argued to be much more precise than other measures, such as %EFC and %EFT (Evans et al. 2014). In the present study, students' writing accuracy was measured by %EFC, %EFT, and WCR.

T-units, clauses, and errors were analyzed by one of the researchers in all of the texts, and the second researcher also rated 12.5% of the texts for the number of T-units, clauses, and errors on the basis of T-unit, clause, and error guidelines, in order to ensure reliability. Interrater reliability was high, above 95% in all cases, so the scores produced by the first rater were used in subsequent analyses. SPSS 22.0 was applied for calculating descriptive and inferential statistics. For each group of Chinese students, repeated measures ANOVAs were used to analyze the differences in their written performance across the different tasks, then independent samples *t*-tests were conducted to compare participants' written performance in each task condition in the two English study contexts (i.e., China versus Hungary). Before running the analyses, we ensured that the assumptions of each statistical test (normal distribution of the variables and the homogeneity of their variance) were met. Besides checking the statistical significance of our results, we also opted for calculating the effect size measure of partial eta squared. Cohen (1988) suggested that partial  $\eta^2$  at 0.01, 0.09, and 0.25 stand for small, medium, and large effect sizes, respectively (Tabachnick and Fidell 2007, p. 55). Thus, we used these guidelines when interpreting our results.

#### 4. Results and Discussion

##### 4.1. Relationships between Task Conditions and the Written Performances of CSH

Table 1 presents descriptive statistics of the performance measures across the different tasks and repeated measures ANOVA test results for CSH. The asterisk after the name of the measure signals a significant difference based on the repeated measures ANOVA test. The repeated measures ANOVA test did not reveal a relationship between fluency, that is, TW and task conditions. However, the large standard deviation of TW in the CT indicated a large variation in this performance measure among these students, showing that while doing this task, some students probably became very excited and engaged and wrote many more words than their peers. This might suggest that producing output on the CT could be especially well-suited for demonstrating the rapid idea generation aspect of creativity (see Albert and Kormos 2004), or it might be particularly enjoyable or engaging for certain students leading to increased productivity: These possibilities should be explored in further studies. The finding that no difference existed between tasks with the TGT and the LGT in terms of the amount of output is inconsistent with some similar studies on oral tasks (e.g., Lambert et al. 2017; Poupore 2014), where learners produced more words in the LGT condition compared to the TGT condition. Our findings suggest that a higher level of learner agency and more potential for creativity in writing tasks do not necessarily lead to more output. The lack of significant differences is also in contrast with predictions of both Robinson's (2001a, 2001b, 2005) Cognition Hypothesis and Skehan's (2001) limited-attentional resources model, as both predict greater fluency in the case of a cognitively less complex task. In this case, this was the TGT as it only involved telling a ready-made story, while the LGT and the CT both necessitated conceptualizing the story as well. This surprising finding might perhaps be a result of the balancing out of agency-related and cognitive complexity-related effects, which were supposed to be acting in opposite directions. Alternatively, on a more general level, they might be linked to the recursive nature of writing tasks that allow for more planning regardless of the particular task type, evening out potential differences present in speaking. When interpreting this



finding from a topic familiarity perspective, since the content of the LGT is expected to be familiar to the learner, the lack of fluency effects seems to be in line with the results of those studies which also failed to demonstrate fluency effects (Kessler et al. 2022; Yang and Kim 2020).

Regarding lexical complexity measured by the D-value and P\_Lex, the repeated measures ANOVA revealed that task conditions were not associated with the D-value. Thus, the texts produced under the different conditions showed similar lexical diversity in terms of the type/token ratio. In contrast, a significant relationship was observed between task conditions and P\_Lex,  $V = 0.32$ ,  $F(2, 18) = 4.28$ ,  $p = 0.03$ , partial  $\eta^2 = 0.32$  with a large effect size. Follow-up comparisons indicated that the pairwise difference between the TGT and the CT was significant ( $p < 0.05$ ). The students produced higher P\_Lex values in the TGT (Mean<sub>TGT</sub> = 1.08, Mean<sub>CT</sub> = 0.77), indicating that more rare and difficult words were used in the TGT condition. A reason for this might be that the predetermined content illustrated by the pictures forced the students to use specific words to express the teacher-selected ideas that were not necessarily easy, while in the CT, the students could decide to use the words they were sure about and avoid more difficult ones, showing that the TGT condition might promote lexical complexity (Kormos 2011; Lambert and Zhang 2019). Another possible explanation for this finding might be that, according to Robinson's (2001b, 2005) Cognition Hypothesis, the CT task was more difficult and complex than the TGT along the resource-dispersing dimension since it required doing two tasks, coming up with a story and telling it, instead of one. Thus, this might have increased the demand on learners' part for conceptualization (Skehan 2001) and planning (Kellogg 1996), as a result of which, students had to devote more attentional resources to content-generating rather than to language form. This finding confirms Robinson's Cognition Hypothesis, which argues that increasing task complexity along the resource-dispersing dimension leads to lower complexity in task performance but also lends support to Skehan's (2001) limited attentional resources model in that increased cognitive complexity is expected to reduce the complexity of task performance, so in this particular case the two models have similar predictions with regard to effects on learners' performance. When examined from a content familiarity perspective, this finding contradicts the results of all previous content familiarity studies since familiar content, which is what students were supposed to use in the LGT, was associated with greater lexical complexity, albeit different measures of it, in all of the studies reviewed here (Bui and Luo 2021; He and Shi 2012; Kessler et al. 2022; Yang and Kim 2020).

Next, looking at measures of syntactic complexity, two of the three ANOVA analyses concerning the subcomponents of syntactic complexity (i.e., general complexity, measured by MLT, and sub-causal complexity, measured by MLC) did not indicate statistically significant differences among the three task conditions. However, a relationship was revealed between task conditions and subordination, measured by MNCT,  $V = 0.43$ ,  $F(2, 18) = 6.78$ ,  $p = 0.01$ , partial  $\eta^2 = 0.43$  with a large effect size. Follow-up comparisons indicated that the pairwise difference between the TGT condition and the LGT condition was statistically significant ( $p < 0.01$ ). The students produced a higher MNCT in the LGT condition (Mean<sub>TGT</sub> = 1.23, Mean<sub>LGT</sub> = 2.43), indicating that the sentences the students wrote in the LGT condition contained more subordination. It seems that complete control over what to write and how to write it in the LGT enabled the students to focus more on the formulation phase of writing, where they not only generated content but also encoded the language for later use (Kellogg 1996). In other words, they might have written more smoothly using more complex sentence structures during the writing process. Since Pallotti (2009) and Abrams (2019) proposed that complexity might be linked to personal preferences, the interpretation that the LGT allows for a way of writing that is more in line with the participants' personal preferences, thus leading to higher complexity, cannot be ruled out either. However, in the TGT, the students might have simply translated the content presented in the pictures, resulting in a simpler narrative structure, which required the use of shorter sentences with simple sentence structure.

**Table 1.** Descriptive statistics and repeated measures ANOVA test results for Chinese students studying in Hungary (CSH).

Measures	N	TGT	LGT	CT	F	p	Partial Eta Squared
		Mean (SD)	Mean (SD)	Mean (SD)			
TW	20	194.10 (57.89)	198.65 (58.87)	242.40 (111.78)	1.42	0.27	0.14
D-value	20	64.70 (12.10)	65.15 (19.06)	59.55 (14.88)	1.35	0.28	0.13
P_Lex *	20	1.08 (0.34)	0.87 (0.43)	0.77 (0.40)	4.28	0.03	0.32
MLT	20	8.94 (1.51)	9.30 (1.84)	8.97 (2.35)	0.25	0.79	0.03
MNCT *	20	1.23 (0.13)	2.43 (1.57)	1.27 (0.19)	6.78	0.01	0.43
MLC	20	7.31 (1.14)	6.87 (1.28)	7.09 (1.73)	0.61	0.56	0.06
%EFC	20	68.25 (21.85)	68.17 (18.37)	58.93 (24.23)	0.13	0.20	0.20
%EFT	20	62.48 (25.55)	65.66 (19.97)	58.93 (24.23)	1.21	0.32	0.12
WCR	20	0.88 (0.09)	0.87 (0.11)	0.87 (0.11)	0.17	0.84	0.02

Note. TGT = Teacher-Generated Content Task; LGT = Learner-Generated Content Task; CT = Creative Task; N = Number; SD = Standard Deviation; TW = Total number of Words; MLT = Mean Length of T-unit; MNCT = Mean Number of Clauses per T-unit; MLC = Mean Length of Clause; %EFT = Error-Free T-unit Ratio; %EFC = Error-Free Clause Ratio; WCR = Weighted Clause Ratio, \* signals  $p < 0.05$ .

Moreover, the students might have been more willing to complete the LGT and might have been more engaged in doing it since they were given more control over the content to write and could write anything of personal interest and personal relevance, which might have aroused their intrinsic interest and helped foster their writing engagement, thus leading to better performance (Maehr 1984). This finding is partially consistent with previous findings regarding the effects of the TGT and the LGT conditions on oral task performance, namely that the LGT condition yielded more complex language (Lambert and Minn 2007; Lambert et al. 2017; Poupore 2014).

Interpreted from the perspective of Robinson's (2001b, 2005) Cognition Hypothesis, the LGT was probably more complex along the resource-dispersing dimension since it involved doing two tasks at the same time. Thus, it should have led to lower syntactic complexity, as it did in the case of lexis. However, being given a chance to produce their own content (context familiarity) might have counterbalanced this effect because the processing load of coming up with their past experiences might not have been so taxing, leaving available resources for the complex syntactic formulation of the complex content. In fact, this finding is consistent with studies on topic familiarity effects by Yoon (2017) and Kessler et al. (2022), where students produced higher syntactic complexity on the more familiar topic task that was relevant to their lives, and learners had more knowledge about, which was the LGT in our study.

Besides this cognitive explanation, it is also possible that the LGT might have exerted its influence via some affective means, for example by evoking positive emotions and task motivation either related to the particular story itself or related to the feeling of greater agency in determining task content. Although the role of affective factors is acknowledged in Robinson's (2001a, 2001b, 2005) Triadic Componential Framework under *learner factors*, there is minimal research on ways they might be associated with actual task performance. Since we have no information about our participants' perceptions concerning their own performance, it is currently impossible to decide whether cognitive (content familiarity) or affective (motivation or positive emotions) could be in the background of these results.

Additionally, the pairwise difference in MNCT between the LGT and the CT was found to be significant ( $p < 0.05$ ), with higher MNCT being produced in the LGT condition ( $\text{Mean}_{\text{LGT}} = 2.43$ ,  $\text{Mean}_{\text{CT}} = 1.27$ ), showing greater use of subordination. Therefore, the potential for creativity in the CT did not seem to necessarily lead to more creative language use or greater complexity, which does not seem to support Maley's (2006) idea that creative writing is likely to foster learners' engagement with language to produce greater complexity. This finding might be explained by Robinson's (2001a, 2005) Triadic Componential Framework, in that the LGT was probably less cognitively complex than the CT because the content to be produced was supposed to be familiar to the students, thus reducing their cognitive load for the conceptualization of the content. In contrast, in the case of the CT, where students needed to create a story out of several words using their imagination, they were not aided by content familiarity and probably needed to devote considerable resources to conceptualization. There is empirical evidence supporting this line of argumentation since tasks with more familiar content did, in fact, result in the production of texts with higher syntactic complexity in certain cases (Kessler et al. 2022; Yoon 2017). Moreover, this finding can also be understood from the perspective of Kellogg's (1996) writing model and Skehan's (2001) limited-attentional resources model. The two tasks differed in the formulation stage of the writing process, which consists of planning ideas, organizing the narrative structure, selecting vocabulary, as well as applying grammar knowledge. In the CT, generating ideas might have taken students more time and cognitive effort than in the LGT, which might have been the reason for lower complexity in the CT.

Regarding accuracy, repeated measures ANOVAs showed that task conditions were not related to any measures of accuracy used in the study in the case of CSH. This might have been due to the participants' intermediate-level English proficiency, which might have affected their text accuracy independently from the effects of task conditions. The students might have interpreted their involvement in the study as a testing situation, and they might have attempted to avoid taking risks of using difficult language so as to avoid making mistakes. Topic/content familiarity seemed to make no significant difference in this regard, which is consistent with the majority of previous studies on topic familiarity effects on accuracy (e.g., Yang and Kim 2020).

#### 4.2. Relationships between Task Conditions and the Written Performances of CSC

Table 2 displays descriptive statistics of the performance measures across the different tasks and repeated measures ANOVA test results for CSC, where the asterisk after a measure signals a significant difference based on the repeated measures ANOVA test. When comparing fluency measured by the number of words produced, the repeated measures ANOVA test revealed a statistically significant difference here,  $V = 0.29$ ,  $F(2, 18) = 3.71$ ,  $p = 0.05$ , partial  $\eta^2 = 0.29$ , unlike in the case of CSH. Follow-up comparisons indicated that the pairwise difference between the TGT and the LGT was significant ( $p < 0.05$ ), with more words being produced in the TGT ( $\text{Mean}_{\text{TGT}} = 166.00$ ,  $\text{Mean}_{\text{LGT}} = 145.20$ ). Although this finding is different from what we found in the case of CSH, it is in line with the predictions of Robinson's (2001a, 2005) Triadic Componential Framework and Skehan's (2001) limited-attentional resources model, both predicting greater fluency in the case of a cognitively less complex task, like the TGT. In the case of the LGT, participants had to spend more time and devote more effort to the planning phase of writing because they needed to recall their past experiences and generate the content, while in the TGT, they did not need to spend time on the planning of the content and on the structure of their narrative, which might have given them more time for more fluent production, resulting in a higher number of words. However, this finding is in contrast with the results of similar studies on the effects of the TGT condition and the LGT condition on oral task performance, which found that students produced more output in the LGT condition (Lambert and Minn 2007; Lambert et al. 2017; Poupore 2014). As we suggested earlier, it is possible that it was precisely this counter-effect that eliminated significant fluency findings in the case of CSH. We propose that maybe in the Chinese context, the opportunity to create their

own content was too new and surprising for the learners, who were unable to avail of this possibility. Thus, increases in fluency due to lower cognitive complexity in the case of the TGT were not counterbalanced by increases in fluency due to having the possibility of creating their own content, characterizing the LGT and the CT. Therefore, the differences in fluency effects might be related to the learning context, more precisely to the lack of experience with personalized and creative content on the learners' part in China. Studies investigating the feature of content familiarity independently of any other task features found increased fluency in the case of tasks with familiar content in a number of cases (Bui and Luo 2021; He and Shi 2012), but in our study, content familiarity is clearly only one of many potentially influential factors.

**Table 2.** Descriptive statistics and repeated measures ANOVA test results for Chinese students studying in China (CSC).

Measures	N	TGT	LGT	CT	F	p	Partial Eta Squared
		Mean (SD)	Mean (SD)	Mean (SD)			
TW *	20	166.00 (48.45)	145.20 (40.00)	157.25 (59.29)	3.71	0.05	0.29
D-value	20	64.50 (14.42)	60.90 (13.80)	59.50 (14.95)	0.78	0.48	0.08
P_Lex	20	0.88 (0.35)	0.88 (0.44)	0.73 (0.30)	1.47	0.26	0.14
MLT	20	8.78 (1.68)	8.51 (1.60)	8.23 (1.36)	1.50	0.25	0.14
MNCT *	20	1.29 (0.13)	2.38 (1.30)	1.29 (0.21)	6.98	0.01	0.44
MLC	20	6.84 (1.18)	6.75 (1.10)	6.47 (1.18)	0.93	0.41	0.09
%EFC	20	56.59 (20.49)	68.23 (18.14)	67.67 (14.37)	2.84	0.08	0.24
%EFT *	20	53.23 (20.10)	68.17 (18.37)	60.40 (17.17)	4.94	0.02	0.36
WCR *	20	0.83 (0.01)	0.87 (0.08)	0.85 (0.07)	4.29	0.03	0.32

Note. \* Signals  $p < 0.05$ .

Regarding lexical complexity, repeated measures ANOVAs showed no relationship between task conditions and the D-value or P\_Lex. Therefore, it seems that for CSC, different task conditions left students' performance regarding lexical complexity unaffected. This goes against the findings of studies on content familiarity, where familiar content was most often associated with greater lexical complexity (Bui and Luo 2021; He and Shi 2012; Yang and Kim 2020). As for syntactic complexity, our results here are the same as what was found in connection with CSH. While the measure of general complexity (MLT) and sub-clausal complexity (MLC) showed no differences across the tasks, repeated measures ANOVA revealed that task conditions were related to the measure of subordination (MNCT),  $V = 0.44$ ,  $F(2, 18) = 6.98$ ,  $p = 0.01$ , partial  $\eta^2 = 0.44$  indicating a large effect size. Follow-up comparisons showed that the pairwise difference between the TGT and the LGT was statistically significant ( $p < 0.01$ ), with lower MNCT being produced in the TGT ( $\text{Mean}_{\text{TGT}} = 1.29$ ,  $\text{Mean}_{\text{LGT}} = 2.38$ ). We can hypothesize the same possible explanations behind this as in the case of CSH: On the one hand, the possible affective effects of having more freedom and control over the contents of the story to write, and on the other, the possible cognitive load reducing feature of the LGT condition linked to content familiarity might be mentioned here, as well. In addition, the pairwise difference between the LGT and the CT was also statistically significant ( $p < 0.01$ ), with higher MNCT being produced in the LGT ( $\text{Mean}_{\text{LGT}} = 2.38$ ,  $\text{Mean}_{\text{CT}} = 1.29$ ), which could again be explained by content

familiarity effects leading to reduced cognitive load in the planning phase, just as in the case of CSH. Support for a link between content familiarity and increased syntactic complexity was found by Kessler et al. (2022) and Yoon (2017), although the measure of syntactic complexity affected was not always the same as in our study.

In terms of accuracy, two out of the three accuracy measures revealed a significant relationship between task conditions and accuracy when tested by the repeated measures ANOVA tests: %EFT,  $V = 0.36$ ,  $F(2, 18) = 4.94$ ,  $p = 0.02$ , partial  $\eta^2 = 0.36$ , and WCR,  $V = 0.32$ ,  $F(2, 18) = 4.29$ ,  $p = 0.03$ , partial  $\eta^2 = 0.32$ . Follow-up comparisons indicated that the pairwise difference between the TGT and the LGT was significant ( $p < 0.05$ ), with higher %EFT being produced in the LGT condition ( $\text{Mean}_{\text{LGT}} = 68.17$ ,  $\text{Mean}_{\text{TGT}} = 53.23$ ) and with higher WCR being produced in the LGT condition ( $\text{Mean}_{\text{LGT}} = 0.87$ ,  $\text{Mean}_{\text{TGT}} = 0.83$ ). In other words, the students seemed to produce greater accuracy in the LGT than in the TGT.

One possible explanation for this might be that the TGT had a predetermined storyline and left little room for the students to adjust the contents to their linguistic resources (Kormos 2011). Moreover, learners probably did not have a chance to purposefully avoid the vocabulary or syntactic constructions that they had not mastered so well yet. Therefore, they produced lower accuracy levels in the TGT. In contrast, in the LGT, the learners were given opportunities to tailor their language to their linguistic resources and avoid the structures they were not sure about, which might have contributed to greater accuracy. Moreover, since the personal story narration is unlikely to have been practiced in their daily English lessons or at English writing classes in China (Chen 2018; Zhang 2012; Zhang 2016), these students might have performed the writing task more carefully and used more attentional resources in expressing their ideas in grammatically accurate English. This finding is consistent with previous findings regarding the learner agency effects on accuracy on oral task performance, where greater accuracy was produced in the LGT condition (e.g., Lambert and Minn 2007; Lambert et al. 2017; Poupore 2014). From the perspective of the topic/content familiarity effects, the finding that the LGT with higher topic/context familiarity elicited greater accuracy is consistent with the results of He and Shi (2012)'s study but contrary to Kessler et al.'s (2022).

#### 4.3. Differences between the Chinese Students Studying in Different Contexts Concerning Their Written Performances in Different Tasks: The Possible Role of the Study Contexts in Written Performances

With regard to the TGT, the independent samples *t*-tests revealed no significant differences between the two groups of students on any of the performance measures analyzed. This might have been due to the task characteristics in the sense that this task required the students to narrate a story that had a tight storyline and clear story structure. Therefore, the students might have just transcribed the information presented by the pictures. Moreover, the students were not likely to expand the story and add more information to it. In this case, it seems that the students did not push themselves to use more complex language to write the story longer or more complex, which seems to be true for both groups. Therefore, we can assume that the English study contexts did not make a difference on the Chinese students' written performances on tasks operated under the TGT condition.

When examining differences between the two learner groups on the LGT, independent samples *t*-tests revealed that there was a statistically significant difference in fluency, indicated by differences in TW ( $p < 0.05$ ) (see Table 3). CSH wrote more than CSC ( $\text{Mean}_{\text{CSH}} = 198.65$ ,  $\text{Mean}_{\text{CSC}} = 145.20$ ). This outcome might be interpreted from a contextual perspective. According to Maehr's (1984) personal investment theory, socio-cultural context and prior personal experiences determine the meaningfulness of educational tasks and situations for a learner, which, in turn, determine learners' willingness to invest their personal resources (i.e., time, talent, and energy) into completing the task. CSC might not have found this genre of English writing very meaningful because in China, performing well at the College Entrance Exam, which is primarily grammatical in focus, is the main goal for learning English in the larger educational context. Therefore, they might not have been very engaged in completing the personal narrative task.



**Table 3.** Comparing the performances of CSH and CSC with regard to TW produced.

Measure	N		Mean		SD		t	df	Sig (2-Tailed)
	CSH	CSC	CSH	CSC	CSH	CSC			
LGT	20	20	198.65	145.20	58.87	40.00	3.36	38	0.02
CT	20	20	242.40	157.25	111.78	59.29	0.30	38	<0.01

Regarding the CT, independent samples *t*-tests showed that there was a significant difference in fluency, indicated by TW ( $p < 0.01$ ), with CSH writing more also in this condition than CSC ( $\text{Mean}_{\text{CSH}} = 242.40$ ,  $\text{Mean}_{\text{CSC}} = 157.25$ ) (see Table 3). This difference might also be explained from a contextual perspective, with reference to task type familiarity. It can be argued that the two groups of students possibly differed in their familiarity with the task types in the current study since their different linguistic, cultural, and social experiences are likely to contribute to the levels of their familiarity with the tasks (He and Shi 2012). To illustrate, CSH might be more familiar with creative writing like the CT and more prepared for it in terms of writing skills since they might have had more experiences with creative writing, especially those students who attend international American or British schools where creative writing is part of their high school English curriculum (Bennett 2019). Although the genres of argumentation and teacher-generated content narration are frequently practiced in high schools in China, other forms of writing are rarely included in the curriculum and therefore are rarely practiced (Chen 2015; Zhang 2016) since they are not tested in the college entrance exam.

## 5. Conclusions and Pedagogical Implications

Our first two research questions referred to differences in language performance that could be related to the different writing tasks, that is the TGT, the LGT, and the CT, which were completed by the two groups of the students examined. A consistent finding in this regard, which emerged both in the case of CSH and CSC, was that greater syntactic complexity, more precisely the use of more subordination, characterized the LGT as opposed to the other two task types. This suggests that more learner agency, thus, more control over task content, was likely to be beneficial for students' written performances in terms of syntactic complexity.

Our other findings appear to be less consistent since they only appeared in the case of one of the learner groups. Therefore, they should be treated with more caution and should be explored in further studies. One such finding is that in the case of CSH, more difficult and rare words were used in the TGT in comparison with the CT, indicating that the task with predetermined content seemed to bring about more lexical complexity than the creative task. Moreover, in the case of CSC, the higher syntactic complexity described above was also accompanied by higher accuracy in the LGT. These positive features were, nevertheless, counterbalanced by more limited output produced on this task, which might be explained by some trade-off effects (Skehan 2001).

When comparing the performance of the two learner groups directly in an attempt to answer our third research question, it was revealed that CSH and CSC only differed significantly in terms of the amount of output on two out of the three tasks. On the one hand, this provided further proof that the two learner groups were indeed comparable despite the proficiency score difference reported earlier, and on the other hand, the differences identified lent themselves nicely to an explanation of the combined effects of task characteristics and study contexts. The fact that the two groups did not differ in the amount of output on the TGT highlights the powerful influence of tight task structure and predetermined content, which is characteristic of these types of tasks. Nevertheless, on the more loosely structured learner-generated content task and creative task, CSC consistently produced less output either because they were unaccustomed to creating their own content,

or they might have judged these tasks irrelevant as no similar tasks are included in their college entrance exam (Chen 2015; Zhang 2012; Zhang 2016).

This brings us to one of the limitations of our study: Since our design was purely quantitative, questions concerning why students solved those tasks the way they did cannot be explained here. Moreover, we have no information about how much the students complied with the instruction, for example, we do not know whether they, in fact, used their personal experiences in the LGT condition or just made up a story using their imagination. A further limitation concerns the sample size, which was relatively small, although the 40 participants produced 120 narratives altogether. Future task-based research on EFL writing is recommended to be conducted on a larger sample, especially when a contextual perspective is taken. Moreover, complementing students' language output with some qualitative data like introspective accounts in the form of retrospective interviews might help to shed light on some of the issues we can currently only hypothesize about. A final limitation concerns the interpretation of our findings: We primarily relied on cognitive complexity models (Robinson 2001a, 2001b, 2005; Skehan 1998, 2001) when interpreting them, but another possibility would have been using recent research on prompt types (e.g., Cho 2019; Huh and Lee 2018; Shi et al. 2020) as a framework of interpretation. Although we think that these studies might offer interesting insights into how learners solve written narrative tasks, we felt that incorporating this perspective was unfeasible in this article.

As regards the wider pedagogical implications of our study with regard to task design and writing instruction, it seems safe to claim that narrative tasks with predetermined content appear to be useful in the sense that learners' output on them is predictable. Moreover, depending on their content, these tasks might be useful in enhancing lexical diversity, pushing learners to experiment with rare lexis. Therefore, despite their somewhat restrictive nature, they should definitely be part of a teacher's repertoire since they can be reliably used with a variety of different students. Their role in enhancing vocabulary development should not be underestimated either: They can serve as perfect tools for practicing and recycling newly learned words. In contrast, learners' performance on tasks where they have more control over task content and besides exercising their autonomy, they might use their creativity as well appears to be influenced by a wider range of factors, including contextual influences, making task output less predictable on them. This is not necessarily a problem, as such tasks have the potential to involve students more and push them beyond their limits in the hope of better performance. What the teacher should keep in mind in connection with such tasks is that certain learners might not be motivated by them, and they might have a hard time availing of the freedom or control such tasks provide for them.

A further lesson to be taken away from our findings is that apart from consistent increases in syntactic complexity on the learner-generated content task, the rest of the findings appeared to be more context-dependent. They suggest that besides the more frequently researched cognitive complexity factors (Robinson 2001a, 2001b, 2005) and considerations of processing constraints (Skehan 2001), other issues like content familiarity concerns should also be taken into consideration in connection with learner-generated content and creative tasks, along with a range of affective factors. Besides the cognitive load reducing effects of content familiarity and the decrease in processing load that this entails, the potentially motivating influence of positive emotional states created by the sense of agency and the opportunity to be creative can be hypothesized to influence learners' performance on these tasks. Similarly, differences in learners' performance in the two learning contexts are also likely to be related to the affective correlates of being able to exercise agency and be creative. While in the ELF context in Hungary, learners probably encountered such tasks before and responded to them positively, which was reflected by increases in fluency. In the Chinese EFL context, the demotivating effects of negative affects linked to the cognitive appraisals of judging such task as irrelevant or less useful might have been at play. Therefore, to maximally engage the students in order to produce optimal performance and enhance language learning, teachers and language practitioners could

gauge students' affective appraisals of the writing tasks and make adjustments in tasks and task choices in order to make them possibly more engaging for learners, especially in the case of students of different cultural and contextual backgrounds. Teachers should make an effort to acknowledge the affective correlates of language learning and tailor their instruction in a way that learners' affective states and learning experiences are optimal. In simple words, teachers should strive to make the learning experience as relevant and meaningful as possible.

Although Robinson's (2001a, 2001b, 2005) Triadic Componential Framework makes a reference to affective factors under the heading of *Learner Factors*, the systematic investigation of the affective determinants of learners' task performance is still a task ahead of us. Despite the fact that more and more SLA researchers agree with the claim that cognition and emotions are inseparable (Swain 2013), the emphasis has clearly been on cognitive factors in task-based research so far. Therefore, it might be time to update and expand our existing frameworks in view of these new insights, as it is quite likely that it is precisely the moderating effects of affective factors that might be in the background of seemingly conflicting results found in different contexts.

**Author Contributions:** Conceptualization, Q.W. and A.A.; Formal analysis, Q.W. and A.A.; Investigation, Q.W.; Methodology, Q.W. and A.A.; Supervision, A.A.; Writing—original draft, Q.W. and A.A.; Writing—review & editing, Q.W. and A.A. All authors have read and agreed to the published version of the manuscript.

**Funding:** The second author is a member of the MTA-ELTE Foreign Language Teaching Research Group, so in her case the writing of the article was funded by the Research Program for Public Education Development of the Hungarian Academy of Sciences.

**Institutional Review Board Statement:** The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of the Doctoral School of Linguistics, Eötvös Loránd University, Budapest (protocol code: BTK/3226/1 and date of approval: 15 October 2018).

**Informed Consent Statement:** Informed consent was obtained from all participants involved in the study.

**Data Availability Statement:** The data from this study can be made available upon request to the corresponding author.

**Acknowledgments:** We would like to thank the head teachers who helped contact the participants and organized for data collection. We would also like to extend our gratitude to the participants of this study.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A. Three Written Narrative Tasks

The Teacher-Generated Content Task (TGT).

Observe the following six pictures carefully and write a story about what happened. You have 30 min. You can write as much as you like, but write minimum 100 words.

A memorable event.

This is something that I will never forget . . . .





### The Learner-Generated Content Task (LGT).

Please, write a story about a memorable event that happened to you. The beginning sentence is given. You have 30 min. You can write as much as you like, but minimum 100 words.

A memorable event.

This is something that I will never forget. . .

The Creative Task (CT).

Create a story, using at least 6 of the following 10 words. You have 30 min. You can write as much as you like, but minimum 100 words.

Words: desert (沙漠), island (岛), museum (博物馆), plane (飞机), blood(血), ceremony (典礼, 仪式), storm (暴风雨), dragon (龙), bomb (炸弹), flying carpet (飞行地毯)

A memorable event.

This is something that I will never forget. . .

## References

- Abdel Latif, Muhammad M. Mahmoud. 2013. What do we mean by writing fluency and how can it be validly measured? *Applied Linguistics* 34: 99–105. [CrossRef]
- Abrams, Zsuzsanna I. 2019. Collaborative writing and text quality in Google Docs. *Language Learning and Technology* 23: 22–42.
- Albert, Agnes, and Judit Kormos. 2004. Creativity and narrative task performance: An exploratory study. *Language Learning* 54: 277–310. [CrossRef]
- Bennett, Colette. 2019. What's in High School English Class? Available online: <https://www.thoughtco.com/creating-a-vertical-coherence-for-english-curriculum-4123868#english-curriculum-and-the-common-core> (accessed on 5 June 2022).
- Bui, Gavin, and Xueya Luo. 2021. Topic familiarity and story continuation in young English as a foreign language learners' writing tasks. *Studies in Second Language Learning and Teaching* 11: 377–400. [CrossRef]
- Chen, Xi. 2018. A Study on the Teaching of Senior High School English Writing and Suggested Solutions to Its Problems-Taking Y City for Example. Master's thesis, Nanjing Normal University, Nanjing, China.
- Chen, Yan. 2015. The Research on Value Orientations of American SAT Essay. Master's thesis, East China University of Science and Technology, Shanghai, China.
- Cho, Minyoung. 2019. The effects of prompts on L2 writing performance and engagement. *Foreign Language Annals* 52: 576–94. [CrossRef]
- Cogo, Alessia, and Jennifer Jenkins. 2010. English as a lingua franca in Europe: A mismatch between policy and practice. *European Journal of Language Policy* 2: 271–94. [CrossRef]
- Cohen, Jacob. 1988. *Statistical Power Analysis for the Behavioral Sciences*. New York: Lawrence Earlbaum Associates.
- DeKeyser, Robert. 2007. *Practice in a Second Language: Perspectives from Applied Linguistics and Cognitive Psychology*. Cambridge: Cambridge University Press.
- Ellis, Rod. 2018. *Reflections on Task-Based Language Learning and Teaching*. Oxford: Oxford University Press.
- Evans, Norman W., K. James Hartshorn, Troy L. Cox, and Teresa Martin De Jel. 2014. Measuring written linguistic accuracy with weighted clause ratios: A question of validity. *Journal of Second Language Writing* 24: 33–50. [CrossRef]
- Foster, Pauline, and Gillian Wigglesworth. 2016. Capturing accuracy in second language performance: The case for a weighted clause ratio. *Annual Review of Applied Linguistics* 36: 98–116. [CrossRef]
- He, Ling, and Ling Shi. 2012. Topical knowledge and ESL writing. *Language Testing* 29: 443–64. [CrossRef]
- Huh, Myung-Hye, and Jongbong Lee. 2018. Task complexity and writing prompts and performance in EFL high school students' narrative writing. *English Teaching* 73: 55–72. [CrossRef]
- Jin, Hongyu, and Tongshun Wang. 2021. Task complexity, working memory capacity and L2 writing performance: The role of learner agency. *Foreign Languages and Their Teaching* 13: 102–13.
- Johnson, Mark 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. [CrossRef]
- Joyce, Caneel K. 2009. *The Blank Page: Effects of Constraint on Creativity*. Berkeley: University of California.
- Kafipour, Reza, Elaheh Mahmoudi, and Laleh Khojasteh. 2018. The effect of task-based language teaching on analytic writing in EFL classrooms. *Cogent Education* 5: 1–16. [CrossRef]
- Kellogg, Ronald T. 1996. A model of working memory in writing. In *The Science of Writing: Theories, Methods, Individual Differences and Applications*. Edited by C. Michael Levy and Sarah Ransdell. Mahwah: Erlbaum, pp. 57–71.
- Kessler, Matt, Wenye Ma, and Ian Solheim. 2022. The effects of topic familiarity on text quality, complexity, accuracy, and fluency: A conceptual replication. *TESOL Quarterly, Early View*. [CrossRef]
- Kim, Hyun-Ju, Stewart Gray, and Christopher Lange. 2021. Instructional Guidance for Promoting Creativity in English as a Foreign Language Writing Classrooms: A Korean Case Study. *RELJ Journal* 1–17. [CrossRef]
- Kormos, Judit. 2011. Speech production and the cognition hypothesis. In *Second Language Task Complexity: Researching the Cognition Hypothesis of Language Learning and Performance*. Edited by Peter Robinson. Philadelphia: John Benjamins, pp. 39–59.
- Kormos, Judit. 2012. The role of individual differences in L2 writing. *Journal of Second Language Writing* 21: 390–403. [CrossRef]
- Kormos, Judit. 2014. *Speech Production and Second Language Acquisition*. New York: Routledge.
- Kormos, Judit, and Anna Trebits. 2012. The role of task complexity, modality, and aptitude in narrative task performance. *Language Learning* 62: 439–72. [CrossRef]
- Kuiken, Folkert, and Ineke Vedder. 2008. Cognitive task complexity and written output in Italian and French as a foreign language. *Journal of Second Language Writing* 17: 48–60. [CrossRef]
- Lambert, Craig, and Danny Minn. 2007. Personal investment in L2 task design and learning: A case study of two Japanese learners of English. *ELIA* 7: 127–48.
- Lambert, Craig, and Grace Zhang. 2019. Engagement in the use of English and Chinese as foreign languages: The role of learner-generated content in instructional task design. *The Modern Language Journal* 103: 391–411. [CrossRef]
- Lambert, Craig, Jenefer Philp, and Sachiko Nakamura. 2017. Learner-generated content and engagement in second language task performance. *Language Teaching Research* 21: 665–80. [CrossRef]
- Larsen-Freeman, Diane. 1978. An ESL index of development. *TESOL Quarterly* 439–48. [CrossRef]



- Lee, Soonyeon. 2009. The Efficacy of Dynamic Written Corrective Feedback on Intermediate-high ESL Learners' Writing Accuracy. Master's thesis, Brigham Young University, Provo, UT, USA.
- Leveldt, Willem J. M. 1993. *Speaking: From Intention to Articulation*. Cambridge: MIT Press.
- Little, Leni Dam, and Lienhard Legenhausen. 2017. *Language Learner Autonomy: Theory, Practice and Research*. Bristol: Multilingual Matters.
- Liu, Yucui. 2020. Gaozhong yingyu xiezuo jiaoxue gaoxiao celue tanjiu [A study on the Senior High school English writing teaching strategies]. *Internet Education* 6: 188.
- Lo, Julia, and Fiona Hyland. 2007. Enhancing students' engagement and motivation in writing: The case of primary students in Hong Kong. *Journal of Second Language Writing* 16: 219–37. [\[CrossRef\]](#)
- Maehr, Martin L. 1984. Meaning and motivation: Toward a theory of personal investment. *Research on Motivation in Education* 1: 115–44.
- Maley, Alan. 2006. Creative writing/reading. In *Creative Writing in EFL/ESL Classrooms II*. Edited by Jayakaran Mukundan. Kuala Lumpur: Pearson Longman.
- Matsumoto, Yumi. 2021. Student self-initiated use of smartphones in multilingual writing classrooms: Making learner agency and multiple involvements visible. *The Modern Language Journal* 105: 142–74. [\[CrossRef\]](#)
- Meara, Paul, and Imma Miralpeix. 2016. *Tools for Researching Vocabulary*. Bristol: Multilingual Matters.
- Michel, Marije. 2017. Complexity, accuracy, and fluency in L2 production. In *The Routledge Handbook of Instructed Second Language Acquisition*. Edited by Shawn Loewen and Masatoshi Sato. New York: Routledge, pp. 50–68.
- Norris, John M., and Lourdes Ortega. 2009. Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics* 30: 555–78. [\[CrossRef\]](#)
- Ong, Justina, and Lawrence Jun Zhang. 2010. Effects of task complexity on the fluency and lexical complexity in EFL students' argumentative writing. *Journal of Second Language Writing* 19: 218–33. [\[CrossRef\]](#)
- Öveges, Enikő, and Kata Csizér. 2018. *Vizsgálat a Köznevelésben Folyó Idegennyelv-Oktatás Kereteiről és Hatékonyságáról [Study on the Frameworks and Effectiveness of Foreign Language Education in Hungary]*. Budapest: Oktatási Hivatal.
- Pallotti, Gabriele. 2009. CAF: Defining, refining and differentiating constructs. *Applied Linguistics* 30: 590–601. [\[CrossRef\]](#)
- Phung, Linh. 2017. Task preference, affective response, and engagement in L2 use in a US university context. *Language Teaching Research* 21: 751–66. [\[CrossRef\]](#)
- Polio, Charlene. 1997. Measures of linguistic accuracy in second language writing research. *Language Learning* 47: 101–43. [\[CrossRef\]](#)
- Polio, Charlene. 2012. The acquisition of second language writing. In *The Routledge Handbook of Second Language Acquisition*. Edited by Susan M. Gass and Alison Mackey. New York: Routledge, pp. 319–34.
- Polio, Charlene, and Mark Cosgrove Shea. 2014. An investigation into current measures of linguistic accuracy in second language writing research. *Journal of Second Language Writing* 26: 10–27. [\[CrossRef\]](#)
- Poupore, Glen. 2014. The influence of content on adult L2 learners' task motivation: An interest theory perspective. *Canadian Journal of Applied Linguistics/Revue Canadienne de Linguistique Appliquée* 17: 69–90.
- Qiu, Xuyan, and Yuen Yi Lo. 2017. Content familiarity, task repetition and Chinese EFL learners' engagement in second language use. *Language Teaching Research* 21: 681–98. [\[CrossRef\]](#)
- Robinson, Peter. 2001a. Task complexity, cognitive resources, and syllabus design: A triadic framework for examining task influences on SLA. In *Cognition and Second Language Instruction*. Edited by Peter Robinson. Cambridge: Cambridge University Press, pp. 287–318.
- Robinson, Peter. 2001b. Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics* 22: 27–57. [\[CrossRef\]](#)
- Robinson, Peter. 2005. Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *International Review of Applied Linguistics in Language Teaching* 1: 1–33. [\[CrossRef\]](#)
- Ruiz-Funes, Marcela. 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. [\[CrossRef\]](#)
- Samuda, Virginia, and Martin Bygate. 2008. *Tasks in Second Language Learning*. Basingstoke: Palgrave Macmillan.
- Segalowitz, Norman, Barbara Freed, Joe Collentine, Barbara Lafford, Nicole Lazar, and Manuel Díaz-Campos. 2004. A comparison of Spanish second language acquisition in two different learning contexts: Study abroad and the domestic classroom. *Frontiers: The Interdisciplinary Journal of Study Abroad* 1: 1–18. [\[CrossRef\]](#)
- Seidlhofer, Barbara. 2005. *English as a Lingua Franca*. Oxford: Oxford University Press.
- Shi, Bibing, Liyan Huang, and Xiaofei Lu. 2020. Effect of prompt type on test-takers' writing performance and writing strategy use in the continuation task. *Language Testing* 37: 361–88. [\[CrossRef\]](#)
- Skehan, Peter. 1998. *A Cognitive Approach to Language Learning*. Oxford: Oxford University Press.
- Skehan, Peter. 2001. Tasks and language performance assessment. In *Researching Pedagogic Tasks: Second Language Learning, Teaching, and Testing*. Edited by Martin Bygate, Skehan Peter and Merrill Swain. New York: Routledge, pp. 167–85.
- Skehan, Peter. 2009. Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics* 30: 510–32. [\[CrossRef\]](#)
- Swain, Merrill. 2013. The inseparability of cognition and emotion in second language learning. *Language Teaching* 46: 195–207. [\[CrossRef\]](#)

- Tabachnick, Barbara G., and Linda S. Fidell. 2007. *Using Multivariate Statistics* 5. Boston: Pearson.
- Tavakoli, Parvaneh. 2014. Storyline complexity and syntactic complexity in writing and speaking tasks. In *Task-Based Language Learning: Insights from and for L2 Writing*. Edited by Heidi Byrnes and Rosa M. Manchón. Amsterdam: John Benjamins Publishing Company, pp. 217–36.
- Tin, Tan Bee. 2011. Language creativity and co-emergence of form and meaning in creative writing tasks. *Applied Linguistics* 32: 215–35. [\[CrossRef\]](#)
- Tin, Tan Bee. 2012. Freedom, constraints and creativity in language learning tasks: New task features. *Innovation in Language Learning and Teaching* 6: 177–86. [\[CrossRef\]](#)
- van Lier, Leo. 2004. *The Ecology and Semiotics of Language Learning: A Sociocultural Perspective*. Dordrecht: Springer Netherlands.
- van Lier, Leo. 2008. Agency in the classroom. In *Sociocultural Theory and the Teaching of Second Languages*. London: Equinox.
- Vygotsky, Lev Semenovich. 1978. *Mind in Society: Development of Higher Psychological Processes*. Cambridge: Harvard University Press.
- Wang, Liping, and Chunlan Jin. 2022. Effects of task complexity on linguistic complexity for sustainable EFL writing skills development. *Sustainability* 14: 4791. [\[CrossRef\]](#)
- Wigglesworth, Gillian, and Pauline Foster. 2008. Measuring Accuracy in Second Language Performance. Paper presented at the Annual Conference of Teachers of English to Speakers of other Languages (TESOL), New York, NY, USA, October 26–28.
- Williams, Jessica. 2012. The potential role(s) of writing in second language development. *Journal of Second Language Writing* 21: 321–31. [\[CrossRef\]](#)
- Wolfe-Quintero, Kate, Shunji Inagaki, and Hae-Young Kim. 1998. *Second Language Development in Writing: Measures of Fluency, Accuracy, and Complexity*. Honolulu: University of Hawaii Press.
- Yang, Weiwei, and Youjin Kim. 2020. The effect of topic familiarity on the complexity, accuracy, and fluency of second language writing. *Applied Linguistics Review* 11: 79–108. [\[CrossRef\]](#)
- Yang, Wenbin. 2020. *Gaokao Bi Shua Juan, Shinian Zhenti (Yingyu) [Must-do Tests for College Entrance Exam, Ten Past Exam Papers (English)]*. Beijing: Kaiming Press.
- Yoon, Hyung-Jo. 2017. Linguistic complexity in L2 writing revisited: Issues of topic, proficiency, and construct multidimensionality. *System* 66: 130–41. [\[CrossRef\]](#)
- Yu, Xianglan. 2017. Gaozhong yingyu xiezuo jiaoxue xianzhuang yu gaijin celve taijiu [A study on current situation and improvements of English writing teaching at Senior High schools]. *Yingyu Jiaoshi* 18: 98–99.
- Yuan, Dan. 2012. Gaozhong yingyu xiezuo jiaoxue de xianzhuang fengxi ji youxiao duice [A study on the analysis of current situation of Senior High school English writing teaching and effective solutions]. *Journal of Yanbian Education College* 26: 3.
- Zhang, Fu. 2012. Gaokao yingyu xiezuo mingti silu jiedu [An interpretation of composition proposition in College Entrance Examination]. *Journal of Teaching and Administration* 31: 72–73.
- Zhang, Xiaodong. 2016. The Composition Analysis of College Entrance Examination between China and America. Master's thesis, Northeast Normal University, Jilin, China.