



Investigation of the Impact of Prompt Type on Writing Performance of Pre-service Teachers based on Continuation Task

Nader Afshounpour¹, Peyman Rajabi^{2*}, Hossein Ahmadi³

¹Ph. D. candidate, Department of English Language Teaching, Malayer Branch, Islamic Azad University, Malayer, Iran

^{2*}Assistant Professor, Department of English Language Teaching, Malayer Branch, Islamic Azad University, Malayer, Iran

³Assistant Professor, Department of English Language Teaching, Malayer Branch, Islamic Azad University, Malayer, Iran

Received: July 13, 2023

Accepted: September 20, 2023

Abstract

This study examined the effect of the writing prompt properties on Iranian EFL pre-service teachers' essay writing. The continuation task, a new form of reading-writing integrated task in which test-takers read an incomplete story and then write the continuation and ending of the story, has been increasingly used in writing assessment. Language-test developers' understanding of the effects of important task-related factors on test-takers' performance with regard to this task is still in its infancy. In this study we study the effect of prompt type on English as a foreign language (EFL) learners' writing performance in a continuation task. Three groups of Iranian EFL learners performed a continuation task with three different prompts and filled out a writing strategy questionnaire. The participants' continuations were scored holistically and textually analyzed using a range of fluency, grammatical accuracy, lexical complexity, syntactic complexity, cohesion, and source-use features. Prompt type significantly affected the participants' overall continuation writing scores, syntactic complexity, cohesion, and source-use features.

Keywords: Writing assessment, Prompt type, Writing performance, EFL writing

INTRODUCTION

Recently, integrated reading-writing tasks have been considerably used in large-scale, high-stakes language tests due to their high degree of authenticity, potential for content-bias reduction, and positive washback (Plakans, 2015; Shin & Ewert, 2014). In the context of Iran, the continuation task, a new form of an integrated reading-writing task that needs learners to read and continue an incomplete story, is increasingly favored in language pedagogy and assessment. The increased attention to this task is based on research evidence of its

facilitative effect on language learning (Wang, 2015). Nevertheless, our understanding of the effects of key task-related factors on test-takers' performance in this task is still in its infancy.

Previous studies on task-related variables in the continuation task focused on the properties of the source text, such as the perceived level of its interest (Xue, 2013) and its degree of linguistic complexity (Peng, Wang, & Lu, 2020). The effect of prompt type, a key consideration in writing task design, has not yet been systematically examined. With the current study, we fill this research gap by investigating the effect of prompt type on test-takers' writing performance and writing strategy use in the continuation

*Corresponding Author's Email:

peymanrajabi25@malayeriau.ac.ir

task. In the rest of this section, we first discuss the theoretical basis of the continuation task and empirical research supporting its language learning potential and its validity and reliability as a writing assessment instrument. We then review previous research on the effect of writing prompts on writing performance.

It can be said that a writing prompt refers to any stimulus provided by a writing task for test-takers to respond to in a writing assessment (Kroll & Reid, 1994). In a narrower sense, it is conceived as a specific set of task instructions or requirements given to test-takers (Wang, 2016).

The writing prompt has been considered a critical task-related variable that may affect test-takers' writing performance, along with other variables such as general task type (e.g., integrated vs. independent writing), genre (e.g., narrative vs. argumentative writing), and topic (e.g., a general vs. specific topic). Theoretically, these variables were posited to elicit differential writing performance from either a task complexity or communicative functional perspective (e.g., Biber, Gray, & Staples, 2016). Research from the task complexity perspective generally sought to support either Robinson's (2001) cognition hypothesis,

Previous studies of the effect of prompt type have examined differences in test-takers' writing scores and the textual features of their responses. Way, Joiner, and Seaman (2000) investigated the effects of three prompts (bare, vocabulary, and prose model) on the writing performance of students across three proficiency levels. They found that the prose model prompt produced the highest mean writing scores and the bare prompt the lowest. Bahrebar and Darabad (2013) reported similar results in their examination of the effects of the same three prompt types on the overall writing quality of Iranian intermediate EFL learners. In terms of textual features, it has been reported that a more explicit prompt tended to elicit more fluent and accurate production (Way et al., 2000), lower syntactic complexity (O'Loughlin & Wigglesworth, 2007), and more varied modes of argumentation (He & Sun, 2015). For the continuation task, previous investigations of prompt effect are

limited to Yuan's (2013) and Xin's (2017) analyses of the effect of explicitness of task instructions on alignment with the source text and target structure use. More research on the prompt effect on test-takers' writing performance in terms of writing scores and a more comprehensive set of textual features will be useful.

Although not a focus in previous studies of prompt type, test-takers' writing strategy use has been examined in previous research on other task-related variables in writing assessment. Through think-aloud protocols and retrospective interviews, Plakans (2008) found that L2 writers performing reading-to-write tasks adopted more discourse synthesis strategies, whereas those performing writing-only tasks experience more initial planning. Through stimulated recall interviews with test-takers, Chapman (2016) found that such prompt characteristics as domain and response mode may affect test-takers' processes and strategies for prompt selection, response planning, and response organization. The effect of prompt type on test-takers' writing strategy use in the continuation task has not yet been examined. In this study we examine the effect of prompt type on Iranian pre-service EFL teachers' writing performance.

METHODS

Participants

The present study was conducted in Khouzestan Farhangian University, Ahvaz city, Iran. 90 participants were chosen from the population of Iranian EFL pre-service teachers majoring in English. They were at the age range of 20 to 24. All of them had prior experience in writing academic English paragraphs. They were taking writing courses to enhance their English writing ability, and they had formal training in English writing skills. The 90 participants selected all scored 12–13 out of a total of 15 points on the pre-writing test as they represented the largest sub-group of the learners. We randomly divided these participants into three groups (30 per group). A one-way ANOVA confirmed that there were no significant between-group differences in the writing scores ($F(3, 116) = .72, p > .05$).

Table 1*Three types of prompt used in the study*

Prompt	Description
P1: Framed prompt	Read the story and continue it based on the opening sentence of each paragraph provided
P2: Vocabulary prompt	Read the story and continue it using at least five key words underlined in the given passage
P3: Bare prompt	Read the story and continue the story to make it complete

Table 2*Some of measures used to analyze test-takers' continuations*

Category	Measures	Tool
Fluency	Text length	Coh-Metrix 3.0
Grammatical accuracy	Three-point grammatical accuracy scale	Human rating
Lexical complexity	Measure of Textual Lexical Diversity	Coh-Metrix 3.0
	Incidence of content words	
	Concreteness of content words	
	Imageability of content words	
Syntactic complexity	Mean sentence length	Coh-Metrix 3.0
	Number of words before the main verb	
	Number of modifiers per noun phrase	
	Passive voice density	
	Syntactic similarity of adjacent sentences	
Cohesion	Incidence of all connectives	Coh-Metrix 3.0
	Incidence of causal connectives	
	Incidence of logical connectives	
	Incidence of adversative and contrastive connectives	
	Incidence of temporal connectives	
	Incidence of additive connectives	
	LSA similarity between adjacent paragraphs	
	LSA similarity between adjacent sentences	
Source use	LSA Given-New	AntConc 3.5.7
	Source-oriented four-word sequences	

The continuation task with three different prompts.

One weekend in July, Jane and her husband, Tom, had driven three hours to camp overnight by a lake in the forest. Unfortunately, on the way an unpleasant subject came up and they started to quarrel. By the time they reached the lake, Jane was so angry that she said to Tom, "I'm going to find a better spot for us to camp" and walked away. With no path to follow, Jane just walked on for quite a long time. After she had climbed to a high place, she turned around, hoping to see the lake. To her surprise, she saw nothing but forest and, far beyond, a snow-capped mountain top. She suddenly realized that she was lost.

"Tom!" she cried. "Help!"

No reply. If only she had not left her mobile phone in that bag with Tom. Jane kept moving, but the farther she walked, the more confused she became. As night was beginning to fall, Jane was so tired that she had to stop for the night. Lying awake in the dark, Jane wanted very much to be with Tom and her family. She wanted to hold him and tell him how much she loved him.

Jane rose at the break of day, hungry and thirsty. She could hear water trickling somewhere at a distance. Quickly she followed the sound to a stream. To her great joy, she also saw some berry bushes. She drank and ate a few berries. Never in her life had she tasted anything better.

Feeling stronger now, Jane began to walk along the stream and hope it would lead her to the lake.

As she picked her way carefully along the stream, Jane heard a helicopter. Is that for me? Unfortunately, the trees made it impossible for people to see her from above. A few minutes later, another helicopter flew overhead. Jane took off her yellow blouse, thinking that she should go to an open area and flag them if they came back again.

Prompt 1: Read the story given above and continue the story. The opening sentence of each paragraph is given below, based on which you are supposed to develop a complete story in at least 150 words.

But no more helicopters came and it was getting dark again. It was daybreak when Jane woke up.

Prompt 2: Read the story given above and continue the story. You are supposed to write at least 150 words and involve at least five key words underlined in the given passage.

Prompt 3: Read the story given above and continue the story to make it complete. You are supposed to write at least 150 words.

RESULTS

We analyzed the continuations using 21 features of fluency, grammatical accuracy, lexical complexity, syntactic complexity, cohesion, and source use (see Table 2). We chose these features because they have been reported to correlate with L2 writing quality (Chapman, 2016; Cumming et al., 2005; Gebril & Plakans, 2013; McNamara, Graesser, McCarthy, & Cai, 2014) and because they were judged to be meaningfully related to the evaluation criteria in the rating scale.

Given that the writing task was timed, we measured writing fluency as the total number of words in each continuation (Gebril & Plakans, 2013). We obtained word counts through Coh-Metrix (McNamara et al., 2014).

Grammatical accuracy is understood as the ability to be free from grammatical errors while using language to communicate (Plakans, Gebril, & Bilki, 2016). Consistent determination and classification of grammatical errors in L2 production have been shown to be challenging

(e.g., Cumming et al., 2005). We opted for a simple holistic measure of grammatical accuracy following Cumming et al. (2005) and Gebril and Plakans (2013). This measure uses a three-point scale to characterize the grammatical accuracy of a writing sample, with 1 indicating many errors (e.g., over three per T-unit, often affecting comprehensibility), 2 some errors (two to three per T-unit; comprehensibility largely unaffected), and 3 few or no errors (comprehensibility unaffected). Grammatical accuracy scoring was independently performed by the same two raters, with an inter-rater reliability index (measured using Pearson's correlation) of .84 ($p < .01$).

Lexical complexity refers to the variation and sophistication of the words in a text (Lu, 2012). Based on previous research on lexical complexity (McNamara et al., 2014; Riazi, 2016) and the narrative nature of the continuation writing task, we adopted the following four measures from Coh-Metrix: (1) Measure of Textual Lexical Diversity (MTLD), a measure of lexical diversity found not to be affected by text length; (2) incidence of content words (i.e., number of nouns, adverbs, adjectives, and main verbs per 1000 words); (3) concreteness of content words, a measure of the extent to which the content words are concrete or abstract; and (4) imageability of content words,² a measure of the ease to construct mental images for the content words.

Syntactic complexity, that is, the degree of sophistication and variation of the structures produced, has been operationalized in many ways (Lu, 2017). We adopted five indices incorporated in Coh-Metrix: (1) mean sentence length; (2) number of words before the main verb; (3) number of modifiers per noun phrase; (4) passive voice density (i.e., number of agentless passive forms per 1000 words); and (5) syntactic similarity, a measure of the extent to which adjacent sentences in a sample have similar structures. A higher value in the first four measures is associated with a higher degree of syntactic sophistication, whereas a higher degree of syntactic similarity is associated with a lower degree of syntactic variation (McNamara et al., 2014).

Cohesion features are explicit characteristics in a text that help create cohesive links between

ideas and clauses (McNamara et al., 2014). We assessed the cohesion of the continuations using six incidence scores and three Latent Semantic Analysis (LSA) indices in Coh-Metrix. The incidence scores were for all connectives, causal connectives (e.g., *because*), logical connectives (e.g., *if*), adversative and contrastive connectives (e.g., *although*), temporal connectives (e.g., *when*), and additive connectives (e.g., *more-over*). The LSA indices were LSA similarity between adjacent paragraphs, LSA similarity between adjacent sentences, and LSA Given-New, which estimates the proportion of new information in each sentence. The LSA indices range from 0 to 1, with a higher value associated with greater cohesion (McNamara et al., 2014; Riazi, 2016).

Source use refers to the extent to which a continuation aligns with the source text, operationalized as the proportion of the top 20 most frequent four-word sequences in the continuation that were source-oriented (i.e., they also appeared

in the source text) (Wang & Wang, 2015). Specifically, following Wang and Wang (2015), we first identified the top 20 most frequent four-word sequences in the continuations in each group using AntConc 3.5.7 (Anthony, 2018), then determined which of the 20 sequences were source-oriented, and finally calculated the source-use ratio as the ratio of the token frequency of the source-oriented sequences to the token frequency of the top 20 most frequent four-word sequences in each group.

In terms of the overall writing scores, Group 3 ($M = 17.41$) and Group 1 ($M = 13.78$) obtained the highest and lowest mean score, respectively (see Figure 1). The Levene test indicated homogeneity of variance among the three groups ($p = .15$). As shown in Table 4, a one-way ANOVA revealed significant between-group differences in the mean scores ($F(3, 116) = 4.92$, $p = .00$, $\eta^2 = .11$). Pairwise comparisons using the LSD post hoc test indicated that Group 1 had a significantly lower mean score than Groups 2 ($p = .02$) and 3 ($p = .01$).

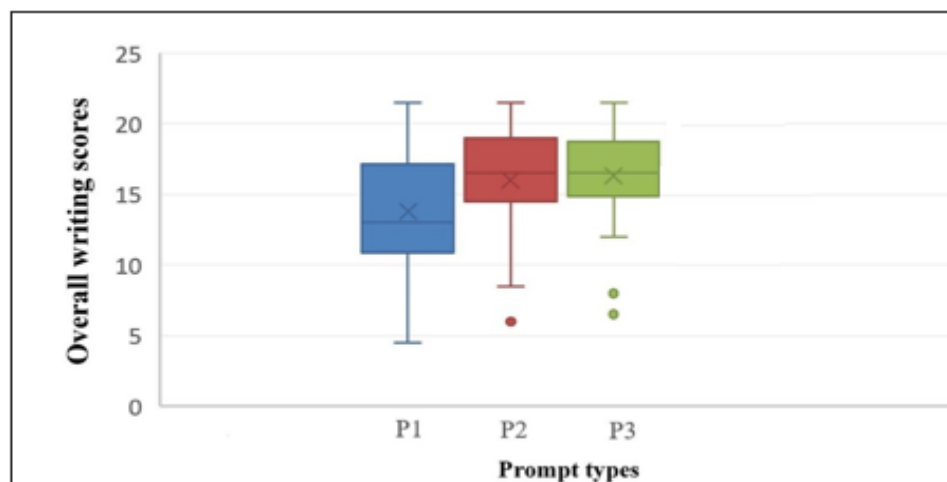


Figure 1

Box-and-whisker plot for overall writing scores for different prompt types. The bottom and top of the box represent the first and third quartiles, the line in the box represents the median, and the x in the box represents the mean. The dots represent outliers: one for P2 (6), two for P3 (6.5 and 8)

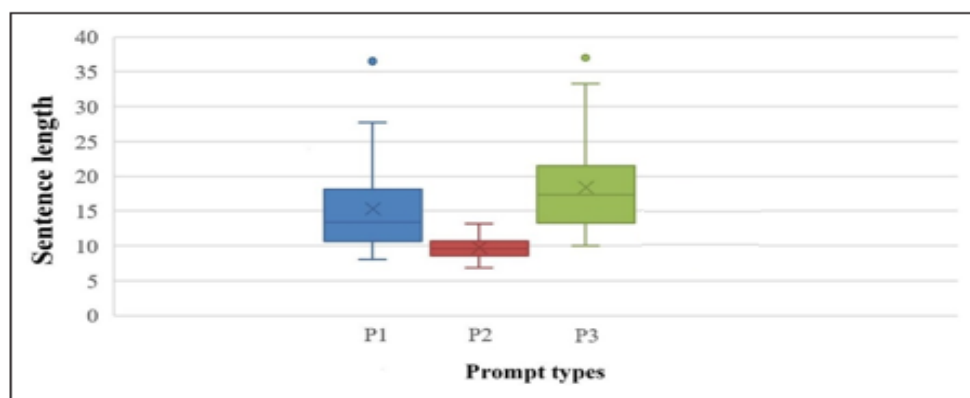


Figure 2

Box-and-whisker plot for sentence length for different prompt types. The bottom and top of the box represent the first and third quartiles, the line in the box represents the median, and the x in the box represents the mean. The dots represent outliers: one for P1 (36) and one for P3 (36.4).

Significant between-group differences were found in four syntactic complexity measures. For mean sentence length ($F(3, 116) = 17.64, p = .00, \eta^2 = .31$) and number of modifiers per noun phrase ($F(3, 116) = 7.30, p = .00, \eta^2 = .16$), Group 3 and Group 2 showed the highest and lowest value, respectively (see Figures 2 and 4). LSD post hoc tests revealed that Group 2 produced significantly shorter sentences and fewer modifiers per noun phrase than all other three groups. For number of words before the main verbs ($F(3, 116) = 5.35, p = .00, \eta^2 = .12$), Group 3 ($M = 3.65$) and Group 2 ($M = 2.42$) also showed the highest and lowest value, respectively (see Figure 3); Games-Howell post hoc tests indicated that Group 2 had a significantly lower value in this measure than Group 3. For syntactic similarity of adjacent sentences ($F(3, 116) = 11.12, p = .00, \eta^2 = .22$), Groups 1 and 3

showed the lowest value ($M = .12$), while Group 2 showed the highest value ($M = .17$) (see Figure 5). In this case, a higher degree of syntactic similarity corresponds to a lower degree of syntactic variation. LSD post hoc tests indicated that Group 2 had a significantly higher degree of syntactic similarity than other groups.

Two of the eight cohesion features showed significant between-group differences, namely, additive connectives ($F(3, 116) = 5.44, p = .00, \eta^2 = .12$) and LSA similarity between adjacent paragraphs ($F(3, 116) = 33.58, p = .00, \eta^2 = .46$) (see Figures 6 and 7). Games-Howell and LSD post hoc tests revealed that Groups 1 and 3 had significantly lower values in both of these cohesion features than Group 2. These results suggest that the continuations produced by Group 2 tended to be more cohesive than those produced by Groups 1 and 3.

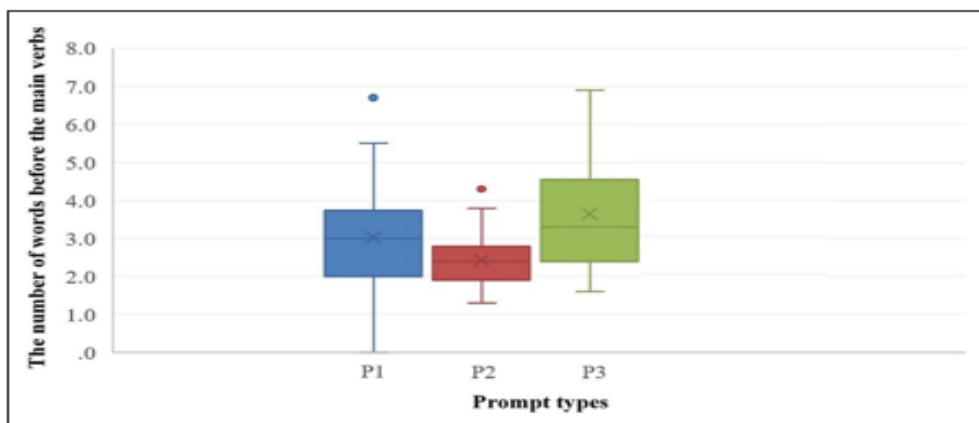


Figure 3

Box-and-whisker plot for the number of words before the main verbs for different prompt types. The bottom and top of the box represent the first and third quartiles, the line in the box represents the median, and the x in the box represents the mean. The dots represent outliers: one for P1 (6.7) and one for P2 (4.3).

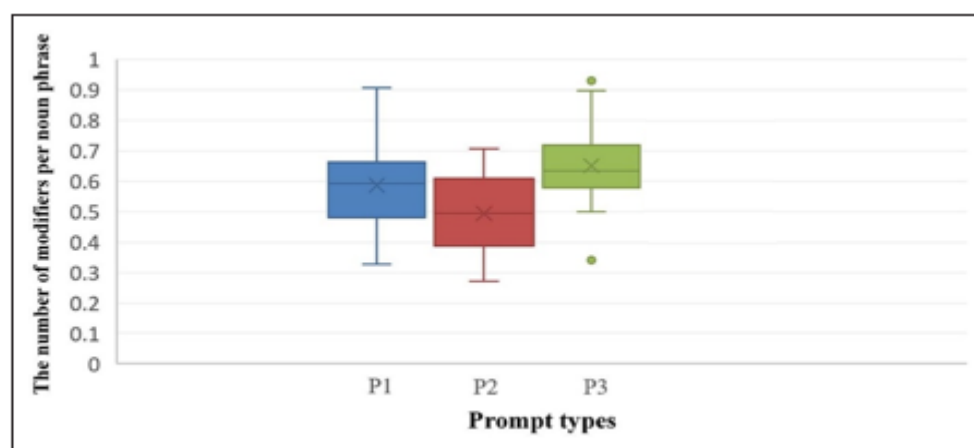


Figure 4

Box-and-whisker plot for the number of modifiers per noun phrase for different prompt types. The bottom and top of the box represent the first and third quartiles, the line in the box represents the median, and the x in the box represents the mean. The dots represent outliers: two for P3 (.92 and .33).

Our results revealed significant effects of prompt type on multiple syntactic complexity, cohesion, and source-use features in the participants' continuations, but not on the fluency, grammatical accuracy, and lexical complexity features considered.

Discussion

In terms of syntactic complexity and cohesion, the framed prompt elicited significantly lower syntactic sophistication and variation than the other prompts, and the framed prompt elicited significantly lower syntactic variation but higher cohesion than the bare prompt and the vocabulary prompt. O'Loughlin and Wigglesworth (2007) reported that prompts with less information tended to elicit more complex language. Our results on the higher syntactic complexity elicited by the bare prompt than by the framed prompt partially support their finding. Our results also suggest that the provision of opening sentences elicited lower syntactic complexity but higher cohesion. This finding may be accounted for using Skehan and Foster's (2001) Limited Attentional Capacity Model, which posited that, due to the limited capacity of human attention, humans must prioritize their attentional resources on tasks with different cognitive demands, resulting in trade-off effects in different areas of performance. When opening sentences for the paragraphs are provided, test-takers must devote attentional resource to ensuring that they develop their story

in a way that logically connects each part to the opening sentence of each paragraph, resulting in higher cohesion. Consequently, less attentional resource can be allocated to the complexity of language, resulting in lower syntactic complexity. The results on the source-use ratio indicated that the prompts with extra elements elicited a higher degree of alignment with the source text than the bare prompt. These results are consistent with Yuan's (2013) finding that explicit prompts may enhance linguistic alignment in L2 writing. Additionally, the provision of required key words resulted in particularly higher source-use ratios. As Wang (2015) and Xiao (2013) noted, L2 writers who referred back to the source text more frequently tended to align more with the source text. The provision of required key words may have prompted the test-takers to refer back to the source text more often, leading to an increased level of alignment.

The three prompts yielded continuations with comparable fluency and grammatical accuracy. These results differ from Way et al.'s (2000) finding that the vocabulary prompt elicited more fluent and accurate writing samples than the bare prompt. This difference may have arisen from the difference in the proficiency level of the participants in the two studies. Whereas Way et al.'s (2000) participants were novice foreign language learners, our participants were at the upper-intermediate level, as evidenced in the scores (12 or 13 out of 15) they earned in the pre-writing test.

CONCLUSION and Implications

With this study we sought to understand the potential effects of prompt type, an important task variable, on test-takers' writing performance and writing strategy use in the continuation task. Our findings revealed that whether the prompt provides opening sentences for the paragraphs or required key words or both may significantly affect test-takers' overall writing scores, their use of syntactic complexity, cohesion and source-use features, and their use of monitoring strategies. Our findings suggest that prompts that include both opening sentences for the paragraphs and required key words are likely to allow test-takers to better demonstrate their full writing ability, compared to prompts that integrate none or one of these elements. Our findings also provide useful information that can inform L2 writing teachers' decision in selecting and designing different writing prompts for the continuation task.

References

- Bahrebar, S., & Darabad, A. M. (2013). Writing in the foreign language classroom: The effects of prompts on the Iranian learners of English. *International Journal of English Language Education*, 1(3), 1–15.
- Biber, D., Gray, B., & Staples, S. (2016). Predicting patterns of grammatical complexity across language exam task types and proficiency levels. *Applied Linguistics*, 37(5), 639–668.
- Chapman, M. (2016). The effect of the prompt on writing product and process: A mixed methods approach (Unpublished doctoral dissertation). University of Bedfordshire, Luton, UK.
- Cumming, A., Kantor, R., Baba, K., Erdosy, U., Eouanzoui, K., & James, M. (2005). Differences in written discourse in independent and integrated prototype tasks for next generation TOEFL. *Assessing Writing*, 10(1), 5–43.
- Gebril, A., & Plakans, L. (2013). Toward a transparent construct of reading-to-write tasks: The interface between discourse features and proficiency. *Language Assessment Quarterly*, 10(1), 9–27.
- He, L., & Sun, Y. (2015). Investigation of the effects of prompt characteristics on Chinese test-takers' integrated writing performance. *Foreign Language Teaching and Research*, 47(2), 237–250.
- Kroll, B., & Reid, J. (1994). Guidelines for designing writing prompts: Clarifications, caveats, and cautions. *Journal of Second Language Writing*, 3(3), 231–255.
- McNamara, D., Graesser, A., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge: Cambridge University Press.
- Peng, J., Wang, C., & Lu, X. (2020). Effect of the linguistic complexity of the input text on alignment, writing fluency, and writing accuracy in the continuation task. *Language Teaching Research*, 24(3).
- Plakans, L. (2015). Integrated second language writing assessment: Why? what? how? *Language & Linguistics Compass*, 9(4), 159–167.
- Plakans, L. (2008). Comparing composing processes in writing-only and reading-to-write test tasks. *Assessing Writing*, 13(2), 111–129.
- Riazi, A. (2016). Comparing writing performance in TOEFL-iBT and academic assignments: An exploration of textual features. *Assessing Writing*, 28, 15–27.
- Shin, S. Y., & Ewert, D. (2014). What accounts for integrated reading-to-write task scores? *Language Testing*, 32.
- Wang, C. (2011). An analysis of three teaching preferences in light of an efficient route to L2 learning. *Foreign Language Teaching and Research*, 43(4), 540–549.
- Wang, C. (2012). The continuation task—An effective method in enhancing foreign language acquisition. *Foreign Language World*, 5, 2–7.
- Wang, C. (2015). Why does the continuation task facilitate L2 learning? *Foreign Language Teaching and Research*, 47(5), 753–762.
- Wang, F., & Zhang, C. (2017). Analysis of take-takers' continuation writing samples in National Matriculation English Test in October 2016. *English Language Learning*, 3, 59–62.

- Way, D., Joiner, E., & Seaman, M. (2000). Writing in the secondary foreign language classroom: The effects of prompts and tasks on novice learners of French. *Modern Language Journal*, 84(2), 171–184.
- Weigle, S. (2002). *Assessing writing*. Cambridge: Cambridge University Press.
- Wilson, M. (1988). MRC psycholinguistic database: Machine-usable dictionary, version 2.00. *Behavior Research Methods, Instruments, & Computers*, 20(1), 6–10.
- Xiao, T. (2013). The effect of alignment on L2 writing accuracy (Unpublished master's thesis). Guangdong University of Foreign Studies, Guangzhou, China.
- Xin, S. (2017). Effects of continuation task condition on L2 grammatical structure acquisition. *Modern Foreign Languages*, 4, 507–517.
- Xue, H. (2013). Interest- related alignment in L2 story continuation (Unpublished master's thesis). Guangdong University of Foreign Studies, Guangzhou, China.
- Yang, W., Lu, X., & Weigle, S. C. (2015). Different topics, different discourse: Relationships among writing topic, measures of syntactic complexity, and judgments of writing quality. *Journal of Second Language Writing*, 28, 53–67.
- Zhang, X. (2016). Multi-dimensional analysis of the effects of the continuation task on the ESL writing process. *Foreign Language World*, 6, 86–94.
- Zhang, X., & Zhou, Y. (2014). *A study on integrated English writing test tasks*. Shanghai: Shanghai University Press.