EFFECT OF CAPTION TYPE IN PICTURE PROMPTS ON EFL WRITING QUALITY

Yali Shi

Jiangnan University, China
E-mail: ylshi@jiangnan.edu.cn

Abstract. *To explore the validity of picture-prompt writing assessment, the study investigated the influence of caption type (i.e. narrative, abstract, and zero) in the picture prompt on EFL writing quality of Chinese college students in terms of holistic scores and functional dimensions (i.e. Dimensions of Involvement, Narration, Elaborated reference, Persuasion, Abstractness, and On-line informational elaboration). ANOVA, MANOVA, and linear regression analysis were conducted and results showed that (a) participants performed significantly better with the abstract caption; (b) Dimensions of Involvement and Abstractness significantly distinguished essays with the abstract caption from those with other types; (c) Dimension of Persuasion significantly predicted ratings of essays with the narrative caption, while Dimension of Narration significantly predicted ratings of essays with the abstract caption. Finally, implications for picture-prompt writing assessment and instruction were discussed.*

Key word: *caption, picture prompt, writing quality, functional dimension, holistic ratings*

1. INTRODUCTION

Writing tasks and their interaction with test takers have been perennial topics in writing assessment research (Hamp-Lyons 2011). Given the more prevalent use of visuals of various kinds in communication in the modern digital age on the one hand and on the other the great value of picture prompts for teaching and assessing writing (Bae & Lee 2010; Olshansky 2018), performance assessment has involved visuals (e.g. pictures, graphs) as prompts to elicit speaking or writing performance, as shown in a number of language tests ranging from large-scale (e.g. IELTS, TOEFL, GSEEE[1], CET[2]) to in-house ones. However, while most studies in the field deal with graph-prompt speaking tasks (e.g. Xi 2010) and writing tasks (e.g. Yang 2016), writing tasks with picture prompts have been inadequately addressed (e.g. Bae & Lee 2010; Li 2018).

[1] Graduate School Entrance English Examination (GSEEE) is administered annually by the National Education Examinations Authority of the Ministry of Education of China. It is high-stakes in nature as passing it is a prerequisite for the entrance into various master programs.
[2] College English Test (CET), administered by the National College English Testing Committee on behalf of the Ministry of Education of China, is a large-scale English test to examine the English proficiency of undergraduates in terms of whether they reach the required English levels specified in the National College English Teaching Syllabuses (Zheng & Cheng 2008).

Thus, in order to explore the validity of picture-prompt writing assessment, as well as to inform development of writing tasks with picture prompts, it is necessary to investigate the influence of picture prompts on writing performance and to identify any possible construct-irrelevant factors.

2. LITERATURE REVIEW

This section presents a review of previous studies on the effect of picture prompts on writing performance, and also introduces the role of caption in picture prompts and multi-dimensional analysis.

2.1. Previous studies on the effect of picture prompts

There are several studies on the effect of picture prompts on writing performance, which display the following three features. First, the picture prompt was usually treated intuitively in terms of the number of frames of pictures (a single picture vs. comic strip) (Bae & Lee 2010), the content of pictures (Jung & Bae 2013; Li 2018), or the order of pictures (ordered vs. disordered) (Kormos 2011). Second, the effect of visual prompts on the quality of written products was mainly examined in terms of ratings (Bae & Lee 2010; Li 2018) or individual linguistic and discourse characteristics, such as lexical complexity, lexical diversity, syntactic complexity, accuracy, cohesive features, and word count (Kormos 2011; Jung & Bae 2013). No significant effect was revealed, except in Kormos (2011)'s study where a series of ordered pictures elicited essays with significantly greater word concreteness and use of connectives. Third, most studies were conducted among EFL learners on primary and secondary levels (Bae & Lee 2010; Kormos 2011; Jung & Bae 2013).

To sum up, previous studies have provided a limited view of the effect of picture prompts on writing performance, both in terms of the independent (i.e., features of the picture prompt) and dependent variables (i.e., quality of writing products). To broaden the view, the following two sections will address these limitations by focusing on the caption in picture prompts and multi-dimensional analysis.

2.2. The caption in picture prompts

The picture prompt in writing tasks is one type of stimulus which refers to "the material that forms the basis for generating writing content" (Weigle 2002: 62). Compared with verbal language, pictures embody distinctive characteristics which pose both benefits and challenges in communication. Specifically, pictures render the conveyed idea more salient owing to their "semantic thickness and sense of presence" (Kjeldsen 2015), yet at the same time, they lead to vagueness and ambiguity of meaning expressed due to their analogical quality and lack of an explicit propositional syntax (Messaris & Abraham 2001). Given these functional and structural differences, pictures should be treated in a different way from verbal language.

To comprehend picture prompts, two levels of processing are involved: the identification of elements in the image and the interpretation of what all these elements are constructed to convey. To facilitate the interpreting process, verbal input is often added, which acts as (a) providing a context to enable a more specific and concrete understanding, (b) facilitating content analysis of images, and (c) stabilizing the broad interpretations of images through

disambiguation or clarification of the meaning (Hagan 2007; Roque 2012). Therefore, with a view to nailing down the exact meaning conveyed by the picture, it is necessary to pay attention to the verbal input (caption) accompanying it, if present.

As to the cognitive processing of the picture prompt with captions, there are two approaches. One is from the angle of cognitive load. According to Cognitive load theory (Sweller et al. 1998), working memory capacity is limited in itself and further constrained by different cognitive loads, one of which is the extraneous cognitive load. It is generated by the manner in which information is presented and thus is under the control of design. During the writing process, writers' working memory could be overburdened, which would lead to failures in any of the stages—planning of ideas, translation of ideas into written sentences, and reviewing the ideas and text already produced (Kellogg 1996). To reduce the overload of working memory, it is necessary to facilitate efficient or even automatic retrieval of relevant content, lexical, grammatical, and discourse knowledge from long-term memory, so that writers can creatively apply what is retrieved to decide what to write and how to write it (Kellogg et al. 2013). In this case, the caption in the picture could function as a means to manipulate the extrinsic cognitive load of the task.

The other is from the angle of cognitive representation. Specifically, in the Conceptual Structuring System (Evans & Green 2006: 526), there is an Attentional System governing the distribution of attention. It is controlled by, among others, how patterns of attention are organized. One of the attention patterns is the Figure-Ground organization, dictating what is fore-grounded and what is back-grounded in cognitive processing, and further determining lexical-grammatical use. In this case, the caption in the picture could act as a means to manipulate the figure-ground attention pattern in processing the prompt.

To find out whether caption in the picture prompt plays a role in writing performance and what role it plays if present, two types of caption were designed for this study together with the zero caption type. One was narrative caption, projecting the thoughts of characters in the picture; the other was abstract caption, adding qualifying information about the meaning of the picture.

2.3. Multi-dimensional analysis (MD analysis)

As mentioned in Section 2.1, previous studies often examined the effect of picture prompts in terms of individual linguistic features of writing products. However, Jarvis et al. (2003) found that writing quality may depend more on linguistic features used together than on individual features. Despite some of the features (e.g. syntactic complexity, lexical diversity, and word frequency) were found significantly predictive of essay quality (e.g. McNamara et al. 2010; Taguchi et al. 2013), these features only revealed the degree rather than the manner of variation among essays, given the fact that the distribution and interrelationship among them were not taken into consideration.

To explore linguistic features used together, MD analysis provides an approach. It is a corpus-driven methodological approach that identifies the frequent linguistic co-occurrence patterns in a language, relying on inductive empirical analysis (Biber 2010: 179). The original Biber's MD analysis (1988) investigated the frequency and distribution of 67 linguistic features of 481 spoken and written texts from 23 sub-registers of the London-Oslo-Bergen Corpus and London-Lund Corpus. Then the normalized frequency counts for each feature (per 1000 words) were obtained and further exposed to factor analysis to reduce the linguistic features to six linguistic co-occurrence patterns (i.e. Dimensions of Involvement, Narration, Elaborated

reference, Persuasion, Abstractness, and On-line informational elaboration), which were interpreted as underlying functional dimensions based on the situational, social, and cognitive functions shared by these co-occurring features. Since these dimensions have both linguistic and functional content, MD analysis has an advantage over the analysis based on individual linguistic features alone.

There have been two orientations for the application of Biber's MD analysis. One is its application to the study of the linguistic characteristics of more specialized registers and discourse domains, such as the university spoken and written registers (Biber et al. 2004; Deroey & Biber 2007), female and male conversational style (Biber & Burges 2000), etc. The other is its application to the investigation of microscopic linguistic variation in (a) L2 writing across such parameters as task types and test-taker proficiency levels (Biber & Gray 2013; Biber et al. 2016), text types and first language groups (Weigle & Friginal 2015), time of writing and essay scores (Friginal & Weigle 2014), as well as contexts of writing (writing in tests vs. in disciplinary learning) (Staples et al. 2018); or (b) in L1 argumentative writing across such parameters as prompts, grade levels, geographical areas, etc. (Crossley et al. 2014). Results of these studies suggested that linguistic co-occurrence data could be used to further characterize students' writing across different parameters. Arguing for the use of MD analysis in language assessment, Biber et al. (2016) held that the dimensions resulted "provide a potentially useful set of holistic measures for testing applications" and these measures are "linguistically well motivated and interpretable". Thus, more MDA of L2 writing products across different parameters (e.g. the prompt) are called for (Friginal & Weigle 2014).

As regards how to conduct MD analysis, there are generally two approaches. One is a full MD analysis that runs an independent exploratory factor analysis to produce new dimensions for a previously unstudied register, and the other makes use of previously established dimensions that have been produced (Conrad & Biber 2001). This study, instead of conducting a new EFA, intends to utilize the six functional dimensions derived from Biber's MD study (1988) for two reasons. First, these dimensions have been treated as general dimensions of variation which are used to characterize various registers. Second, a new MD analysis would be less useful for an analysis of a single register (Biber et al. 2004).

2.4. Research questions

Taken together, this study aims to investigate the influence of caption type (i.e. narrative, abstract, and zero) in the picture prompt on EFL writing quality, as shown by functional dimensions (i.e. linguistic co-occurrence patterns) as well as holistic scores. Specifically, the study addresses the following three research questions:

1. Does the writing quality vary significantly across three types of captions in terms of holistic ratings?
2. Does the writing quality vary significantly across caption types in terms of functional dimensions?
3. What is the possible relationship between functional dimensions and holistic ratings across caption types?

3. METHODS

3.1. Participants and materials

Participants in this study were 154 juniors from three parallel classes of the education major in a Chinese college. The original picture-prompt writing task[3] consists of the instruction and a picture as the prompt without any verbal input in it. Students are asked to write an essay of 160-200 words within 35 minutes in which they need to describe the picture briefly, explain the intended meaning, and give comments. The picture prompt depicts a football match in which a goal keeper is standing in front of a goal, ready to protect it, while a striker is trying to kick the ball in. The goal keeper is imaging in his mind that he is so tiny compared with the size of the goal, while the striker is visualizing in his mind that the goalkeeper is so giant as almost to intercept the goal. The picture prompt intends to convey such an idea that self-confidence is essential for people to achieve success.

There were three types of caption in the picture prompt for this study: the narrative caption, the abstract caption, and the zero caption (see Appendix 1). Specifically, the narrative caption projected the thoughts of both the goalkeeper and the striker in the verbal mode by means of two bubbles with words in the picture, thus providing test takers with a clear knowledge of the two characters' mental activities; the abstract caption "*the psychology on the football field*" went beyond what is presented explicitly in the picture and revealed in a general way the implied meaning of the picture; the zero caption meant that no verbal input was added to the original picture prompt, so test takers had to make their own interpretations based on the picture alone. In all, they varied in the amount of aid for test takers to interpret the picture.

3.2. Data collection

The picture-prompt writing task was administered during regular class period. The between-subjects design was used in that the participants of the three classes were randomly assigned one of the three versions of the writing task. It was found by ANOVA that there was no significant difference in writing scores of the final examination of English last semester among the three groups (F (2,151) =.016, p =.984 >.05), suggesting that any possible difference in writing quality could not result from writing proficiency. Table 1 presents the number of participants, the total word count, as well as the word count per essay for each caption type.

Table 1 Distribution of participants across different captions

Caption type	Narrative caption	Abstract caption	Zero caption
Num. of participants	52	55	47
Total words	8404	10099	7948
words per essay	162	184	169

As to essay rating, two trained raters scored each essay based on a 0-20 holistic scoring rubric (see Appendix 2). The Pearson correlation between the two raters' ratings was .83. The averages of the two ratings were used as final marks of participants' writing. If the

---

[3] It is selected from picture-prompt writing tasks of GSEEE.

scores given by two raters were over 2 points away from each other, another rater would be involved as an adjudicator, and then the average of the two closer scores among the three would be used as the final mark.

3.3. Data analysis

First, to explore whether writing quality varied significantly across caption types in terms of holistic ratings (RQ1), one-way ANOVA was used. Second, to answer RQ2 (i.e. whether writing quality varied significantly across caption types in terms of functional dimensions), first of all, Multidimensional Analysis Tagger[4] (MAT) 1.3 (Nini 2015) was used to calculate the normalized frequency of 67 linguistic features and the scores on six functional dimensions. Next, MANOVA was used to check whether these six dimension scores varied significantly. If there was a significant effect, then six univariate ANOVA would be conducted for each dimension with the alpha level adjusted by Bonferroni correction (i.e. .05/6 = .008). And for post-hoc comparisons in each ANOVA, the alpha level would be further adjusted (i.e. .008/3 = .003). Finally, to find out what was the possible relationship between functional dimensions and holistic ratings across caption types (RQ 3), linear regression analysis was conducted to calculate the predictive power of six dimension scores for the ratings across caption types.

4. RESULTS

4.1. Influence of caption type on writing quality in terms of holistic ratings

As to the effect of different caption types on scores, it was shown by ANOVA that different caption types significantly influenced writing scores $(F (2,151) = 7.705, p = .001, R^2 = .093)$. Post-hoc LSD results revealed that performance with the abstract caption ($M = 10.2, SD = 3.951$) was significantly higher than that with the narrative caption ($M = 7.692, SD = 3.444$) ($p = .002$) and the zero caption ($M = 7.723, SD = 3.916$) ($p = .006$) respectively. But the performance with the narrative caption and the zero caption did not differ from each other significantly.

4.2. Influence of caption type on writing quality in terms of functional dimensions

Table 2 presents both descriptive and inferential statistics of six dimension scores across three caption types.

In terms of central tendency, scores for Dimensions of Involvement, Narration, and On-line informational elaboration were the highest in essays with the narrative caption than in those with the other two captions, while scores for Dimensions of Elaborated reference, Persuasion, and Abstractness were the highest in essays with the abstract caption than in those with the other two captions. As regards dispersion, scores for Dimensions of Involvement, Narration, and On-line informational elaboration spread most widely in essays with the narrative caption than in those with the other two dimensions; for essays with the

---

[4] MAT is a computer program that replicates Biber's (1988) tagger for the MD analysis of English texts. It not only generates a grammatically annotated version of the texts together with the statistics needed to perform MD analysis but also plots the texts on Biber's (1988) six dimensions.

zero caption, scores of Dimensions of Persuasion and Abstractness spread most widely than for essays with the other two captions; scores of the Dimension of Elaborated reference spread most widely in essays with the abstract caption than in those with the other two captions.

Table 2 Functional dimension scores across three types of captions

| Dimension | Caption | | | | | | ANOVA | | |
| | Narrative | | Abstract | | Zero | | | | |
	M	SD	M	SD	M	SD	F	p	R^2
1 Involvement	12.50	9.42	3.60	6.46	11.43	8.87	18.296	**.000**	.195
2 Narration	-2.39	2.98	-3.74	2.36	-3.40	2.73	3.577	.030	.045
3 Elaborated reference	1.64	3.21	3.34	3.71	1.89	3.47	3.731	.026	.047
4 Persuasion	2.74	4.59	2.90	4.77	2.32	5.37	.185	.832	.002
5 Abstractness	-0.64	3.23	3.87	2.99	0.22	3.65	28.333	**.000**	.273
6 On-line informational elaboration	0.55	2.88	0.50	2.58	0.05	2.25	.553	.576	.007

MANOVA results indicated that using Wilks's lambda, there was a significant effect of caption type on scores of six functional dimensions ($L = 0.577$, $F (12, 292) = 7.687$, $p = .000$, $R^2 =.240$). The follow-up ANOVA results revealed that the caption only significantly influenced scores of Dimensions of Involvement ($F (2,151) =18.296$, $p =.000 < .008$, $R^2 =.195$) and Abstractness ($F (2,151) = 28.333$, $p =.000 < .008$, $R^2 = .273$), as shown by the number in bold print in Table 3.

Post-hoc LSD results showed that scores for the Dimension of Involvement (Dimension 1) with the abstract caption were significantly lower than that with the narrative caption ($p = .000 < .003$) and the zero caption ($p = .000 < .003$) respectively. To be specific, essays with the abstract caption exhibited relatively more prominent co-occurring use of linguistic features with negative loadings on Dimension 1, such as attributive adjective and nouns. In contrast, essays with the narrative caption had relatively more prominent co-occurring use of linguistic features with positive loadings on Dimension 1, such as emphatics (e.g. just, such a, really, so, a lot, for sure), second person pronouns (e.g. you, your, yours), and *be* as main verb (e.g. is, are, am). And essays with the zero caption had relatively more prominent co-occurring use of such positive features on Dimension 1 as first person pronouns (e.g. I, me, we, our, ourselves, us), private verbs (e.g. think, believe, find, feel, fear, guess, show, realize), and subordinator *that*-deletion. To interpret the linguistic difference, essays with the narrative caption and the zero caption were more interactional and affectively involved, whereas those with the abstract caption were less interactional with some prominent information-oriented features.

On the other hand, scores for the Dimension of Abstractness (Dimension 5) with the abstract caption were significantly higher than that with the narrative caption ($p = .000 < .003$) and the zero caption ($p = .000 < .003$) respectively. Specifically, such linguistic features as conjuncts (e.g., furthermore, thus, however, for example, in other words, as a result), adverbial subordinators (e.g., because, if, unless, though, while, as long as), and *by*-passives were comparatively more frequently used in essays with the abstract caption. Put another way, essays with the abstract caption were relatively more abstract and thus more impersonal and formal than those with the other two captions. To sum up this section, the

picture prompt with the abstract caption elicited significantly less interactional but more abstract essays than that with either narrative or zero caption.

4.3. Relationship between functional dimensions and holistic ratings across caption types

As to how six functional dimension scores and holistic scores were related, only two significant correlations were revealed by Pearson correlation analysis. Specifically, one was the positive correlation between Persuasion Dimension (Dimension 4) scores and scores of essays with the narrative caption ($r = .278, p = .046 < .05$), suggesting that the more frequently such features as infinitives, prediction modals, persuasive verbs, conditional subordination, necessity modals, split auxiliaries, and possibility modals were used together in an essay, the more highly it would be rated. The other was the positive correlation between Narration Dimension (Dimension 2) scores and scores of essays with the abstract caption ($r = .340, p = .011 < .05$), suggesting the more frequently such features as past tense verbs, third person pronouns, perfect aspect verbs, public verbs, synthetic negation, and present participial clauses were used together in an essay, the more highly it would be rated.

Further analysis through multiple linear regression of these two significant correlations generated two significant results. On the one hand, for essays with the narrative caption, scores of the Persuasion Dimension (Dimension 4) significantly predicted scores ($\beta = 0.278 t = 2.046, p = .046$), accounting for 7.9% of their variance ($F (1, 50) = 4.272, p < .05, r^2 = .079$). On the other, for essays with the abstract caption, scores of the Narration Dimension (Dimension 2) significantly predicted scores ($\beta = 0.340, t = 2.630, p = .011$), accounting for 11.5% of their variance ($F (1, 53) = 6.921, p < .05, r^2 = .115$).

5. DISCUSSION

5.1. Influence of caption type on writing quality in terms of holistic ratings

It was found that participants performed significantly better with the abstract caption than with either the narrative or the zero caption respectively. There are both contrary and similar findings in previous studies. For instance, it was found that such factors as the number of frames and the content of the picture prompt did not exert a significant effect on writing performance (e.g. Bae & Lee 2010; Li 2018). Yet, a series of disordered pictures was found to generate significantly poorer quality in terms of word concreteness and cohesion than a series of ordered pictures (Kormos 2010), which was attributed to the differential cognitive complexity of the tasks. Likewise, the findings in this case could be related with the cognitive load (Sweller et al. 1998) involved in processing the picture prompt with different types of captions. Specifically, in the case of picture-prompt writing tasks, test takers are asked to describe the picture, interpret its meaning, and then to give comments. All three interactive sub-tasks must be processed in the working memory for the fulfillment of the requirement of the whole writing task, resulting in a high intrinsic cognitive load, especially the sub-task of interpreting the meaning of the picture. In terms of efficient retrieval from long-term memory of the relevant content and linguistic knowledge for the generation of ideas and sentences, the abstract caption could possibly play a relatively more facilitative role. Specifically, it provided a direct bridge to the implied meaning of the picture, thus more readily activating test-takers' schemas and

reducing the working memory demand of composing online especially in the time-constrained testing situation. In contrast, the narrative caption only re-represented the explicit meaning of the picture, whereas the picture prompt without caption provided no clue to its implied meaning. In these two cases, much of the working memory capacity would be devoted to hesitating about the theme or composing. As found by Kellogg et al. (2013), lack of accessibility of relevant knowledge in working memory contributes significantly to problems in coherence, grammar, and spelling. Therefore, with the help of the liberated working memory capacity for formulation, translation, and reviewing, it could be possible that essays with the abstract caption displayed relatively better quality in general.

5.2. Influence of caption on writing quality in terms of functional dimensions

It was found that Dimensions of Involvement and Abstractness significantly distinguished essays with the abstract caption from those with the other two caption types, suggesting essays with the abstract caption were more information-oriented and abstract. This could be related with the Figure-Ground attention pattern (Evans & Green 2006). In this case, as the abstract caption provided the clue to the implied meaning of the picture, it would most probably be fore-grounded in cognitive processing, whereas the picture itself would most probably be back-grounded. As a result, test takers would concentrate on the abstract caption and elaborate on the implied meaning of the picture in a logical and objective way; and correspondingly more linguistic features (e.g. by-passives, conjuncts, and adverbial subordinators) would be used together, which would add to the abstractness and information-orientation of essays. In contrast, with the narrative or the zero caption, the picture prompt (together with the narrative caption) would possibly be fore-grounded, and test takers would possibly focus on describing the mental activities of the characters in the picture or the picture itself, which would lead to more narrative and less abstract essays in either case. This may suggest that the picture prompt with the caption stating the explicit meaning of the picture (i.e. the narrative caption) would be no more useful than that without caption in eliciting information-loaded and abstract essays. Yet, the above tentative explanation requires verification by test-takers' think-aloud protocols about the writing process.

In all, the findings in this regard corroborate to some extent the argument that patterns of linguistic co-occurrence compensate for the limitation of individual linguistic features and can be used to further analyze writing quality (e.g. Jarvis et al. 2003; Friginal et al. 2014). Specifically, the results echo the previous findings that functional dimensions can reflect the variations in writing products by different task features (e.g. Crossley et al. 2014; Biber & Gray 2013; Weigle & Friginal 2015; Biber et al. 2016; Staples et al. 2018).

5.3. Relationship between functional dimensions and holistic ratings across caption types

Functional dimensions were found to be significantly predictive of essay ratings. Specifically, with the narrative caption, the more overly persuasive the essays were, the higher their ratings were; with the abstract caption, the more narration-oriented the essays were, the higher their ratings were. However, it was the functional dimension concerning the difference between written and spoken language (i.e. formal vs. informal register) that was found in previous studies to significantly distinguish highly-rated essays from lowly-rated ones (e.g. Friginal & Weigle 2014; Biber et al. 2016).

Nevertheless, the results are generally in line with earlier findings that linguistic features used together better captured writing quality than separate features (e.g. Jarvis et al. 2003; Biber & Gray 2013; Biber et al. 2016). Therefore, the findings of this study add further proof to the "linguistically motivated" and "parsimonious" nature of functional dimensions (Biber et al. 2016) and their applicability to the analysis of task-induced variation in essay quality.

6. CONCLUSION

The present study was the first attempt to investigate the influence of caption type in the picture prompt on writing quality. The study revealed that caption types resulted in significant variation in holistic ratings and functional dimensions. And it further indicated the significantly predictive power of two functional dimensions (i.e. Dimensions of Narration and Persuasion) for holistic ratings of essays with two different caption types (i.e. abstract and narrative ones).

In the light of the findings, there are two implications for picture-prompt writing assessment and instruction as follows. First, for assessment, given the significant influence of caption type on picture-prompt writing quality, it is necessary to control for this factor in test design. For example, test developers could optimize the cognitive load of picture interpretation by providing the abstract caption which gives the clue to the implied meaning of the picture prompt. Therefore, the full play of writing proficiency would be neither hindered by the difficulty in picture comprehension nor disturbed by the caption which only re-represents the explicit content of the picture.

Second, for instruction, instead of leaving students reciting writing templates to cope with the test, it is much more beneficial to train students how to flexibly allocate working memory capacity and coordinate three processes of writing based on the cognitive load of picture-prompt writing tasks. For example, to reduce working memory overload in writing, it is necessary to develop students' prewriting strategies such as writing outlines. Besides, in learning from model essays, teachers could remind students to note co-occurring linguistic features instead of separate features and to reflect on their possible communicative functions. More importantly, students should be made aware that the key to good writing lies in how to coordinate linguistic material of various types for specific writing purposes.

Finally, the findings of this study must be treated with caution. The current study is limited in its focus on one kind of picture and topic area in the prompt as well as a limited number of participants from one college. Future studies could examine other picture types or topic areas and involve participants of a wider range in different contexts, so as to further explore how writing quality is conditioned by caption type in picture-prompt writing tasks and to accumulate more validity evidence for picture-prompt writing assessment.

REFERENCES

Bae, J., & Lee, Y. S. 2010. "The validation of parallel test forms: 'mountain' and 'beach' picture series for assessment of language skills." *Language Testing,* 28(2), 155-177.

Biber, D. 2010. "Corpus-based and corpus-driven analyses of language variation and use." In *The Oxford handbook of linguistic analysis*, edited by N. F. Heine & H. Narrog, 159-191. Oxford University Press.

Biber, D., & Burges, J. 2000. "Historical change in the language use of women and men." *Journal of English Linguistics,* 28(1), 21-37.

Biber, D., Conrad, S., Reppen, R. Byrd, P., Helt, M., Clark, V., & Urzua, A. 2004. *Representing Language Use in the University: Analysis of the TOEFFL 2000 Spoken and Written Academic Language Corpus.* TOEFL Monograph Series (MS-25). Princeton, NJ: Educational Testing Service.

Biber, D., & Gray, B. 2013. *Discourse characteristics of writing and speaking task types on the TOEFL iBT1 Test: A lexico-grammatical analysis.* ETS Research Report RR-13-04, TOEFLiBT-19. Princeton, NJ: Educational Testing Service.

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.

Conrad, S., & Biber, D. (Eds.). 2001. *Multi-dimensional studies of register variation in English.* Harlow, England: Pearson Education.

Crossley, S. A, Allen, L. K., & McNamara, D. S. 2014. "A Multi-Dimensional analysis of essay writing: What linguistic features tell us about situational parameters and the effects of language functions on judgments of quality." In *Multi-dimensional Analysis, 25 years on: A Tribute to Douglas Biber*, edited by Sardinha, T. B., & Pinto, M. V., 197-237. J. Benjamins.

Deroey, K. & Biber, D. 2007. "University language: A corpus-based study of spoken and written registers." *Applied Linguistics,* 28(4): 624-627.

Evans, V., & Green, M. 2006. *Cognitive Linguistics: An Introduction.* Edinburgh: Edinburgh University Press.

Friginal, E., Li, M., & Weigle, S. C. 2014. "Revisiting multiple profiles of learner compositions: A comparison of highly rated NS and NNS essays." *Journal of Second Language Writing*, 23, 1-16.

Friginal, E., & Weigle, S. 2014. "Exploring multiple profiles of l2 writing using multi-dimensional analysis." *Journal of Second Language Writing*, 26, 80-95.

Hagan, S. M. 2007. "Visual/Verbal collaboration in print: Complementary differences, necessary ties, and an untapped rhetorical opportunity." *Written Communication,* 24(1), 49-73.

Hamp-Lyons, L. 2011. "Writing assessment: shifting issues, new tools, enduring questions." *Assessing Writing*, 16(1), 3-5.

Jarvis, S., Grant, L., Bikowski, D., & Ferris, D. 2003. "Exploring multiple profiles of highly rated learner compositions." *Journal of Second Language Writing*, 12(4)*,* 377–403.

Jung, J., & Bae, J. 2013. "The influence of picture prompt variation on writing performance: 'Series' vs. 'Imagine Before and After'." *English Language Teaching*, 25(2), 27- 46.

Kellogg, R. T. 1996. "A model of working memory in writing." In *The science of writing: Theories, methods, individual differences, and applications*, edited by Levy, C. M., & Ransdell, S., 57-71. Mahwah, NJ: Lawrence Erlbaum Associates.

Kellogg, R. T., Whiteford, A. P., Turner, C. E., Cahill, M., & Merlens, A. 2013. "Working Memory in Written Composition: An Evaluation of the 1996 Model." *Journal of Writing Research*, 5(2), 159-190.

Kjeldsen, J. E. 2015. "The rhetoric of thick representation: how pictures render the importance and strength of an argument salient." *Argumentation,* 29(2), 197-215.

Kormos, J. 2011. "Task Complexity and Linguistic and Discourse Features of Narrative Writing Performance." *Journal of Second Language Writing*, 20(2), 148-161.

Li, J. 2018. "Establishing Comparability across Writing Tasks with Picture Prompts of Three Alternate Tests." *Language Assessment Quarterly,* 15(4), 368-386.

McNamara, D., Crossley, S.A., & McCarthy, P. 2010. "Linguistic features of writing quality." *Written Communication,* 27(1), 57-86.

Messaris, P., & Abraham, L. 2001. "The role of images in framing news stories." In *Framing public life: Perspectives on media and our understanding of the social world*, edited by Reese, S. D. et al., 215-226. Mahwah, NJ: Lawrence Erlbaum Associates.

Nini, A. 2015. *Multidimensional Analysis Tagger (Version 1.3).* Available at: http://sites.google.com/site/multidimensionaltagger.

Olshansky, B. 2018. "The universal language of pictures: A critical tool for advancing student writing." *TESOL Journal,* 9(4), 1-16.

Roque, G. 2012. "Visual argumentation: A further reappraisal." In *Topical Themes in Argumentation Theory: 20 Exploratory Studies*, edited by van Eemeren, F. H., & Garssan, B., 273-288. Springer Netherlands.

Staples, S., Biber, D., & Reppen, R. 2018. "Using Corpus-Based Register Analysis to Explore the Authenticity of High-Stakes Language Exams: A Register Comparison of TOEFL iBT and Disciplinary Writing Tasks." *Modern Language Journal,* 102(2), 310-332.

Sweller, J., Van Merrienboer, J. J., & Paas, F. G. 1998. "Cognitive architecture and instructional design." *Educational psychology review,* 10(3), 251-296.

Taguchi, N., Crawford, W., & Wetzel, D. Z. 2013. "What linguistic features are indicative of writing quality? A case of argumentative essays in a college composition program." *TESOL Quarterly,* 47(2), 420-430.

Weigle, S. C. 2002. *Assessing Writing*. New York: Cambridge University Press.

Weigle, S. C., & Friginal, E. 2015. "Linguistic dimensions of impromptu test essays compared with successful student disciplinary writing: Effects of language background, topic, and L2 proficiency." *Journal of English for Academic Purposes*, 18, 25-39.

Xi, X. 2010. "Aspects of performance on line graph description tasks: Influenced by graph familiarity and different task features." *Language Testing*, 27(1), 73-100.

Yang, H. C. 2016. "Describing and interpreting graphs: the relationships between undergraduate writer characteristics and academic graph writing performance." *Assessing Writing*, 28, 28-42.

Zheng, Y., & Cheng, L. 2008. "College English Test (CET) in China." *Language Testing*, 25(3), 408-417.