Assessing Writing 38 (2018) 10–20

Contents lists available at ScienceDirect

Assessing Writing

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

Connecting writing assessment with critical thinking: An exploratory study of alternative rhetorical functions and objects of enquiry in writing prompts

Fulan Liu[a,*], Paul Stapleton[b]

[a] Foreign Languages College, Jiangxi Normal University, 99 Ziyang Road, Nanchang, Jiangxi, 330022 China
[b] Department of English Language Education, The Education University of Hong Kong, 10 Lo Ping Road, Tai Po, New Territories, Hong Kong, China

ARTICLE INFO

Keywords:
Writing prompt
Rhetorical function
Object of enquiry
Writing performance
Critical thinking

ABSTRACT

This exploratory study investigated the effects of prompts on students' writing performance and critical thinking. Two groups of Chinese undergraduates (n = 129) responded to two distinct prompts with differing rhetorical functions and objects of enquiry. The comparison group responded to a conventional prompt modelled after prompts in high-stakes English tests which tend to converge over narrow ranges, while the experimental group responded to a prompt with problem-solving as the main rhetorical function, and behavioral economics as the object of enquiry. Various differences were noted between the two sets of essays related to standardized indexes of writing quality, as well as other rhetorical and linguistic features including: use of metadiscourse, essay organization, and use of certain lexical items. The comparison group outperformed the experimental one in several aspects of writing. However, the experimental group generated more diverse language and displayed enhanced critical thinking. We argue that the washback of high-stakes tests may result in a better performance in some respects in students' responses to conventional prompts. However, this may also reflect a cookie-cutter response to such prompts that arrives at the expense of a whole range of ignored rhetorical functions and objects of enquiry, some of which may stimulate increased critical thinking.

1. Introduction

Writing prompts are essential elements in both the assessment and pedagogy of English writing (Crusan, 2010; Weigle, 2002). They have thus become the focus of investigation in recent years. Rhetorical functions and objects of enquiry are two of the key variables when analyzing a writing prompt. Rhetorical function refers to the discoursal purpose of the written response required by a prompt, e.g., evaluation, recommendation or explanation, while object of enquiry refers to the topic of the expected response. (Example prompts from past high-stakes tests are listed in Table 1 to illustrate the two terms). The literature indicates that both the rhetorical function and object of enquiry of a writing prompt may have significant effects on linguistic and rhetorical features within a response as well as the overall quality of students' writing (e.g., He & Shi, 2012; Hinkel, 2009; Lee & Anderson, 2007). However, to date few studies have investigated prompt effects on students' critical thinking despite the close interaction between critical thinking and writing. Critical thinking is defined as a set of cognitive abilities and thinking dispositions (Facione, 2015; Halpern, 2003)[1] which

* Corresponding author.
E-mail addresses: fulan.liu@yahoo.com (F. Liu), paulstapleton@gmail.com (P. Stapleton).

[1] For the purposes of the present study, cognitive abilities are the focus while the dispositional aspects of critical thinking have been left aside.

https://doi.org/10.1016/j.asw.2018.09.001
Received 8 November 2017; Received in revised form 2 September 2018; Accepted 10 September 2018

Table 1
Example prompts for rhetorical functions and objects of enquiry.

Example Prompt	Rhetorical Function	Object of Enquiry
Using a computer every day can have more negative than positive effects on your children. *Do you agree or disagree? (IELTS)*	Evaluation (Making a judgment about the value or importance of a given entity or phenomenon)	The effects of using a computer on a daily basis
Do you agree or disagree with the following statement? *For successful development of a country, a government should focus its budget more on very young children education rather than on universities. (TOEFL)*	Hortation (Making an appeal about some actions as to whether they should be taken or not)	The focus of a country's educational budget
Illiteracy has traditionally been viewed as largely a third world problem. However it is becoming apparent that in countries such as the USA and Australia, illiteracy is on the increase. *Discuss possible causes for this and its effect on society. (IELTS)*	Explanation (Giving an account of the causes for a given entity or phenomenon)	The possible causes and effects of illiteracy

This table is adapted from Authors (2015).

are primarily embodied in people's language (oral or written arguments) (Palmer, 2012; Ramage, Bean, & Johnson, 2010). And according to Flower and Hayes (1981); MacArthur and Graham (2016) and many other researchers, the writing process involves intensive cognitive activities. These cognitive abilities (critical thinking) and cognitive activities (writing process) eventually manifest themselves in a writer's rhetoric. On the theoretical and empirical basis stated above, we think that by analyzing the linguistic and rhetorical features in students' essays, we would be able to investigate their critical thinking.

In addition to possible immediate effects on students' writing performance, writing prompts in high-stakes tests may have salient washback on writing pedagogy (Green, 2006; Qi, 2007). Green (2006), for instance, observed that a considerable amount of time in the writing classroom was spent on practicing prompts conforming to the IELTS (a high-stakes English test) standards. As such, appropriate writing assessment can be used to induce intended educational purposes, such as cultivating critical thinking abilities or other essential skills in students. However, studies (Liu & Stapleton, 2015; Moore & Morton, 2005) show that both rhetorical function and object of enquiry in the writing prompts of high-stakes tests may converge over narrow ranges. Specifically, rhetorical function tends to require evaluation or hortation (strong appeal or recommendation) and object of enquiry tends to be sociocultural in nature. Moore and Morton (2005) claim that such writing prompts emphasize the "spontaneous expression of opinion" (p. 64), and might bring about formulaic structures and limited sets of expressions and ideas in students' responses. In addition, such prompts may cause corresponding washback in the writing classroom, i.e., teachers may emphasize the practice of certain types of writing over others.

In the present study, the generally uniform writing prompts (in terms of rhetorical function and object of inquiry) found in high-stakes English tests were deemed conventional, based on the results from both Moore and Morton (2005) and Liu and Stapleton (2015), and it is hypothesized that writing prompts with alternative objects of enquiry and rhetorical functions may contribute to more diversity in terms of linguistic/rhetorical features and enhanced critical thinking in students' essays.

To test the hypothesis, we designed an exploratory study involving two groups of undergraduate students from a Chinese university. The study is meant to provide possible directions towards cultivating critical thinking in the English writing classroom. The comparison group responded to a conventional writing prompt; in other words, the prompt was similar to those found in high-stakes English tests. The experimental group responded to a prompt whose object of enquiry and rhetorical function was quite unlike those found in conventional prompts. First, the object of enquiry was in an area other than education, technology or social issues, i.e., the usual areas covered in high-stakes tests. Second, the rhetorical function required more analyzing and problem-solving instead of the usual evaluation and hortation. Scripts were collected and changed into electronic data. Both corpus and content analysis were performed on the data to compare and uncover patterns.

2. Review of literature

2.1. The effect of rhetorical function and object of enquiry in writing prompts

Studies have shown that varying objects of enquiry, i.e., personal or impersonal, specific or general, in writing prompts can elicit varying responses (2009, He & Shi, 2012; Hinkel, 1995; Tedick, 1990; Yu, 2010). Yu (2010), for example, by examining the correlation between the object of enquiry and the lexical diversity in test-takers' essays, found that lexical diversity is higher in test-takers' responses to impersonal topics. Hinkel (1995, 2009) found that students' language use when responding to particular writing prompts is dependent on their cultural and personal background, and topic knowledge. In light of these findings, Hinkel (2009) proposed that the use of new objects of enquiry in prompts that are less proximate to test-takers' own experience and opinions may lead to richer linguistic variety. Since language use reflects one's critical thinking (Palmer, 2012), Hinkel's recommendation (2009) has given impetus to the design of this research project, specifically, the need to explore the possible influence of writing prompts with distinct objects of enquiry on students' written argumentation and critical thinking.

As indicated by the literature, different rhetorical functions in writing prompts may produce various linguistic features (e.g., Hinkel, 2002; Reid, 1990; Spaan, 1993). Reid (1990), for example, compared writers' responses to two types of prompts (in terms of their rhetorical functions): 1) comparison/contrast (C/C) and take a position, and 2) describe and interpret a chart or graph (G). Student writers were native speakers of Arabic, Chinese, Spanish and English. Findings revealed that different rhetorical functions elicited responses that were measurably different in linguistic features. Specifically, writers used significantly more content words

when responding to the C/C prompts than for the G prompts, while they used significantly longer words and significantly more pronouns for the G prompts. However, when responding to the two types of prompts, student writers' syntax did not change. In an earlier study, Durst (1987) investigated the text structures and other linguistic features in the analytical and summary writing of students to compare the writing processes employed by the students in both types of essays. It was found that students used more varied and complex thinking skills in analytical writing. The author, therefore, suggested the use of analytical writing as a heuristic for critical thinking. Durst (1987) study provided additional evidence of the effect of rhetorical function (in that study, analysis vs. summary) on linguistic and rhetorical features of students' writing. Another significant implication of Durst's study (1987) to the present one is the discussion of the connection between rhetorical function and critical thinking via examining linguistic features of student essays.

It has been found that there is a tendency for the rhetorical functions of writing prompts on high-stakes tests to converge, especially when compared with university writing assignments for content subjects. In one study, Moore and Morton (2005) compared IELTS writing prompts with a corpus of writing prompts used in two Australian universities. They examined four elements in the two sets of prompts, namely, genre, information source, rhetorical function and object of enquiry. The study revealed very important differences between the writing prompts of IELTS and those used in university. Specifically, university assignments covered a wider range of rhetorical functions including evaluation (67%), description (49%), summarization (35%), comparison (35%) and explanation (28%), while IELTS prompts were mainly restricted to evaluation (100%) and hortation (70%). They also identified distinct features of IELTS prompts regarding object of enquiry. Notably, they found that despite seemingly diverse categories, IELTS prompts focused on "the responsibilities of certain authorities [that are] closely connected to the rhetorical function of hortation" (p. 63). These findings to some extent inspire further investigation into the narrow range of rhetorical functions and objects of enquiry in writing prompts along with the impact this might have on the writing classroom.

In another study more closely related to this one, Liu and Stapleton (2015) analyzed 120 prompts from three high-stakes English tests, namely, TOEFL, IELTS and TEM4 (Test for English Majors Band 4, a high-stakes test in China), in order to see whether these prompts displayed certain patterns. It was found that evaluation and hortation were the two most common rhetorical functions. As for object of enquiry, notable similarities were found among the three tests with three themes of a sociocultural nature – education, technology and social issues – dominating. These findings thus draw attention to the impact that writing assessment may be having on English writing pedagogy; in particular, there is a concern that writing instruction may be confined to a narrow range of rhetorical functions and objects of enquiry leading to a constricted way of writing and even reasoning. For instance, due to the tendency of prompt uniformity in high-stakes tests, memorized templates may be misused or overused in some writing classrooms to prepare for the tests. As a result, students' writing may involve less original thought and personalized rhetoric than it potentially could.

2.2. Washback of high-stakes writing tests on English writing pedagogy

Studies (e.g., Beck & Jeffery, 2007; Cheng, 1998; Green, 2006) have been conducted to investigate both the beneficial and detrimental consequences of high-stakes writing assessment. Green (2006), for example, in order to explore the washback of the academic writing section of IELTS, investigated the relationship between writing test design features and pedagogical practices by observing the behavior of 197 learners and 20 teachers over 51 classroom hours. He found that IELTS prompts featured: 1) impersonal topics, 2) a limited range in terms of rhetorical function, and 3) expected responses based on personal opinions. To examine how washback of the IELTS writing prompt predicted classroom activities, Green (2006) compared the writing-focused IELTS preparation classes with English for Academic Purposes (EAP) classes at U.K. universities. Classroom observation data showed that 58% and 48% of the time in IELTS preparation classes and in EAP classes respectively was spent practicing impersonal topics conforming to the IELTS standards. In addition, topics of study were comparable in both IELTS and EAP classes. Most notably, little time was spent on personal topics or academic topics in either IELTS preparation classes or EAP classes, although these topics were discussed in more depth in some EAP classes. Evidence in Green's study suggests there was considerable washback of IELTS writing tasks on writing pedagogy conducted in the classes he viewed.

Similarly, in a semester-long study involving interviews with teachers and students, classroom observations, and student writing, Ketter and Pool (2001) investigated the impact of a high-stakes test on a high school writing class. They found that the teachers stressed preparation for a high-stakes test over locally developed standards for assessing writing. This also reduced teachers' engagement in reflective practice that would normally cater for the needs of individual students. In addition, "the criteria for passing the test failed to take into consideration the rich variety of American culture and the complexity of literacy learning" (Ketter & Pool, 2001, p.344).

The aforementioned literature indicates that writing assessment, especially that of a high-stakes nature, can be an active agent for educational change. In recent years, increasing attention has been paid to turning this powerful impact of assessment into an advantage in the classroom (e.g., Andrews, 2004; Shohamy, 2007). What should be highlighted here is that if tests are "properly conceived and implemented," then aligning teaching with what is assessed can facilitate positive educational purposes (Popham, 1987, p. 679).

2.3. Can critical thinking competence be enhanced via writing prompts?

In view of the salient effect of writing assessment on pedagogy and students' writing performance, one question to be raised in this section is: can critical thinking cultivation be encouraged via writing assessment?

Critical thinking, a key component of higher-order thinking, is defined as both a set of cognitive abilities and thinking

dispositions. Critical thinking abilities refer to skills of interpretation, analysis, inference, evaluation, explanation, self-regulation, and problem-solving, etc., and a person with critical thinking dispositions would be apt to use these skills (Facione, 2015; Halpern, 2003). Because critical thinking development is often stated as a primary goal in syllabi for students at various levels worldwide, numerous endeavours have been made towards the cultivation of students' critical thinking in classroom teaching. However, few studies have focused on the potential of using assessment for enhancing critical thinking.

While in a more general sense, Alderson and Wall (1993) argue that "tests are held to be powerful determiners of what happens in the classroom" (p. 115), Eckstein and Noah (1993) more specifically contend that one crucial function of tests is to "[encourage] higher levels of competence and knowledge" (p.11). If the active engagement of critical thinking is included in language tests, teachers and learners are more likely to include practice to this end. Following this line of thought, Shirkhani and Fahim (2011) propose that English as a Foreign Language (EFL) learners' critical thinking abilities be enhanced through assessment practice by outlining some specific suggestions for test designers and English language teachers. And although there is the concern about simply adding another dimension to "teaching to the test," If we use tests with clarified instructional targets, teachers can focus their classroom efforts on getting students to master what they're supposed to learn" (Popham, 2001, p. 20).

One example of an effort to assess students' critical thinking in their writing is a study by Condon and Kelly-Riley (2004). Another is Carroll's study (2007), which examined the "intellectual changes in student writing" (p. 216) during a critical thinking course; these intellectual changes were investigated by measuring differences in varied linguistic variables and cognitive variables. The studies by Condon and Kelly-Riley (2004) and Carroll (2007) suggest that critical thinking can be evaluated in student writing by noting the use of language. Thus, language can be an indicator of critical thinking (Carroll, 2004; Palmer, 2012). A critical thinker knows the "power of loaded language" (Carroll, 2004, p. 29), e.g., hedges (Hyland, 2005), or words with emotive content (Carroll, 2004). And in turn, the use of the loaded language indicates the critical thinking of a person. A critical thinker knows how to use words for persuading others to accept ideas or values. She can use expressions such as "I think", "I feel" to indicate that the ideas expressed are personal opinions rather than facts, and use hedges such as "might" to reduce the strengths of a claim. Although differences in the content of writing prompts can induce variation in students' written language, which in turn reflects their critical thinking competence, to date few empirical studies have been conducted to explore the possibility of enhanced critical thinking competence through the use of different writing prompts.

One study (Furtak & Ruiz-Primo, 2008), although not particularly concerned with writing assessment and critical thinking, merits attention because it investigated the use of prompts in generating students' conceptions about an entity or issue in both written responses and classroom discussions. The researchers investigated the usefulness of four types of formative assessment prompts in eliciting middle school students' ideas. Findings revealed that "prompts with fewer constraints and unfamiliar settings elicited a range of student conceptions in writing" (Furtak & Ruiz-Primo, 2008, p. 799). This study demonstrated the effectiveness of prompts in stimulating higher-order thinking in students' writing. Our study, however, takes the exploration a step further to examine the critical thinking skills employed by students in responding to two types of prompts with distinct content and tasks.

2.4. Research question

The current study thus aims to compare the effect of two types of writing prompts on students' critical thinking ability via analyzing the rhetorical and linguistic features of students' essays. It seeks to answer two research questions:

1 In what ways do the essays of students responding to a "conventional" writing prompt differ from those of an exploratory, non-conventional prompt?
2 Can a non-conventional prompt elicit written responses that display elements of critical thinking?

3. Methods

3.1. Participants

One hundred and twenty-nine first year undergraduate students ranging in age from 18 to 21 participated in the study. They were all English majors studying at a provincial public university in Mainland China. At the time of the study, they were taking an English writing course that met once a week for two hours all with the same instructor. Although they were from different parts of the country, they all met the admission standards of the program, which required a minimum grade of 120 (out of 150) on the National College Entrance Exam of the English subject. All the students had been learning English (as a foreign language) for about 10 years by the time they participated the study. Given their similarity in educational background and English proficiency, they could be regarded as a homogeneous group. As these students were randomly assigned to two classes, one class was designated the comparison group and the other the experimental.

3.2. The writing prompts for both groups

The comparison group (n = 64) was assigned the following prompt specially developed for this study to be in line with the

rhetorical functions and objects of enquiry of the prompts from the essay writing tasks of high-stakes English tests[2]. In response to the prompt, students needed to first evaluate the impact of urbanization; then state their own opinion and support it. As such, the rhetorical functions involved in this writing prompt were mainly evaluative and the object of enquiry was a social issue (urbanization). In addition, the prompt contained a goal instruction similar to those found in most of the essay prompts of high-stakes tests.

Prompt for the comparison group

> China has recently experienced a great expansion of the population and size of its cities. This social process is called urbanization. A recent study by the Asian Development Bank and the National Development and Reform Commission estimates that cities in China will grow by about 15 million people each year and by a total of 230 million over the next 15 years. Do you think urbanization is making people's lives better?
> You will be allowed 35 min. Please write on the answer sheet a composition of about 200 words on the following topic:
> Is urbanization making our life better?

The prompt for the experimental group (n = 65) was informed by the field of behavioral economics. Briefly, behavioral economics focuses on areas of human behavior that help explain the irrationality of economic decisions that standard economics fails to explain. In other words, it explores other domains, such as emotion and shortsightedness that tend to play a significant role in decision-making (Ariely, 2008). We chose this field because of its rich focus on both intuition and logic, as well as its potential for creative and critical thinking. For example, in the prompt below, students need to explain the logic behind a counterintuitive occurrence. Here, one could argue that economics is a specialized field unsuitable for a writing prompt; however, the same could be said for socio-cultural prompts that stray into sociology, technology and education, equally specialized fields. In fact, behavioral economics explores questions about everyday human behavior and decision-making that cross disciplines.

The exploratory prompt was devised so that it would meet the conditions recommended in the literature for prompt content, namely that it be contextualized, authentic, and accessible. For instance, we modified an original problem taken from the field of behavioral economics and situated it in Shanghai using local referents, such as currency, so that it would seem more contextualized and authentic to the participants (who were all Chinese). In addition, to enhance accessibility, we wrote the prompt in language within the comprehension level of the students.

However, more importantly, we also endeavored to create a context that would stimulate original and critical thought, unencumbered by concerns over the participants' uneven schema, by supplying an object of enquiry that none of the students were likely to have been exposed to before. With regard to critical thinking skills, compared with the comparison group prompt, the exploratory prompt specifically required them to solve a conundrum, i.e., use a different set of cognitive skills. In essence, the chosen prompt attempted to move beyond the "take a stance on a socio-cultural issue" type of task to new ground that required more than simply giving reasons to support a point, which in some cases could be a memory task retrieved from previous practice tests (He, 2010). In effect, in devising a new kind of prompt, the goal was to explore the type of language and organization the participants used in order to see if it differed from the comparison group. An equally important goal was to investigate the type of problem solving and connection-making that participants would generate that may be absent in "take a stand" essays because of concerns about prescriptive responses or hackneyed arguments in the public domain.

Prompt for the experimental group (Education Bureau, 2014)

> A day care center in Shanghai had a problem. Although parents were supposed to pick up their children each day by 4 PM, often they would arrive late. This meant that some children and teachers had to wait, causing stress for both. In order to try to solve this problem, two economists offered their help. They created a plan that charged parents a 20 RMB penalty each time they arrived later than ten minutes after 4 o'clock. This penalty was added to the parents' monthly fee of 1500 RMB. Soon after the new system started, however, a strange thing happened. The number of late pick-ups actually increased. In the end, the number of parents arriving late more than doubled. Instead of improving the situation, the economists' plan made things worse. [3] Explain why you think the plan had the reverse effect and how you would change it. You will be allowed 35 min. Please write on the answer sheet a composition of about 200 words.

The two prompts were designed to be comparable in terms of the difficulty of task, the length of required response, and the goal instruction; they were also designed to be similar to those used in standardized tests[4]. However, the two prompts feature distinct rhetorical functions and objects of enquiry. While the experimental group's prompt required respondents to interpret behavior and solve a problem as its rhetorical purpose in the field of economics, the comparison group's prompt fell under typical categories, i.e., evaluation and hortation as the rhetorical function and "social issue" as object of enquiry.

[2] The writing sections of both international high-stakes English tests, e.g., TOEFL, IELTS, and Chinese high-stakes English tests, e.g., Test for English Majors Band 4 (TEM4) tend to comprise two writing tasks: one requires an essay-like response to give and support opinions about a controversial topic and the other requires a non-essay-like response to describe graphs or summarize opposing views. Usually the essay writing task accounts for a larger proportion of the total score, and hence the prompt for the comparison group was modelled after the essay writing prompt.

[3] This prompt was based on, and modified from a study in Haifa, Israel, appearing in the following paper: Gneezy and Rustichini (2000). A fine is a price. *Journal of Legal Studies, 29*, 1-17.

[4] A survey was conducted among a group of 31 students other than the experimental and comparison groups as to the accessibility of the content areas of the two prompts prior to the experiment. 27 students (87%) thought that the two prompts were parallel or roughly parallel in terms of the difficulty of the task. Twenty-eight students (90%) thought that the two prompts were comparable or roughly comparable in terms of the familiarity of the topic. 30 students (97%) considered them as standardized prompts.

Table 2
Metadiscourse classification and examples.

Category	Function	Examples
Textual metadiscourse		
Transition markers	express semantic relation between main clauses	in addition / but / therefore / thus / and
Frame markers	explicitly refer to discourse acts or text stages	firstly/secondly/ finally / all in all/ to conclude
Code glosses	help readers grasp meanings of ideational material	namely / for example / in other words / such as
Interpersonal metadiscourse		
Hedges	withhold writer's full commitment to statements	might / perhaps / it is possible / about
Attitude markers	express writer's attitude to propositional content	in my opinion / I think /I agree

3.3. Data analysis

The scripts of the writing test were collected from both the experimental and comparison groups and each was assigned a code. All the data were then converted to electronic text to facilitate data analysis, resulting in two corpora of over 30,000 words, with the experimental corpus being 14,874 words and the comparison corpus 15,596 words. In view of the exploratory and open-ended nature of the research question, we adopted a data-driven approach with an aim to trace meaningful differences in both sets of responses to the two prompts. Data analysis comprised two phases. In the first phase, preliminary text analysis was performed via an automated text analyzer for general linguistic features of the data, and then further text analysis was conducted for syntactic complexity, lexical diversity and word frequency which are considered proxies for the most predictive linguistic elements of argumentative writing quality (Biber, Gray, & Poonpon, 2011; Lu, 2011; McNamara, Crossley, & McCarthy, 2010). Specifically, McNamara et al. (2010) claim that higher quality essays are likely to be characterized by having "more complex sentences, a greater diversity of words, and less frequent words" (p.63). Although researchers tend to agree that syntactic complexity, lexical diversity and word frequency are indicators of writing quality, there are various measures for each of these indicators. In this study, following McNamara et al. (2010), and using the computer program Coh-Metrix 3.0, the syntactic complexity was measured by the number of words before the main verb; lexical diversity was measured by the Measure of Textual Lexical Diversity (MTLD) which has been found to be a reliable index of a writer's vocabulary and indicative of linguistic skills (McNamara et al., 2010); word frequency was measured by Celex (log frequency of all words).

In the second phase, qualitative and corpus analyses were performed focused on: 1) metadiscourse, 2) essay structure, and 3) lexical items. We analyzed the use of metadiscourse in students' scripts for a better understanding of the linguistic features in writing. According to Hyland (2005), metadiscourse is a key element of discourse analysis since it provides "insights into patterns of interaction and engagement" and reveals "how writers, through their texts, see the values, interests and assumptions of their communities" (p.195). Following Crismore, Markkanen, and Steffensen, (1993) and Hyland (2005), two categories of metadiscourse were coded: textual metadiscourse and interpersonal metadiscourse. Under the textual metadiscourse category, we focused on three subcategories: transition markers, frame markers and code glosses. Under the interpersonal categories, two subcategories were coded: hedges and attitude markers. These five subcategories were chosen because the initial coding indicated that these were the most frequent. Table 2 illustrates the coding scheme for metadiscourse use in both the experimental and comparison group essays. The first author coded the metadiscourse in all the scripts (N = 129). A second coder, with a master's degree in applied linguistic with an appropriate research background, coded 50% of the scripts. The inter-rater reliability, calculated with Cohen's kappa, was 0.85, 0.89, 0.90, 0.87, and 0.81 for transition markers, frame markers, code glosses, hedges, and attitude markers respectively. Disagreements were then resolved through discussion. Upon completing the coding of metadiscourse use and data entry, descriptive statistics were generated; then independent sample *T*-tests were performed to compare the use of the five subcategories of metadiscourse between the two groups.

The organizational structure of the essays was also compared by noting the number of paragraphs in each script and analyzing the first three sentences of the first paragraph and comparing the two groups. Customarily, a writer begins an argumentative essay with a preamble of a few sentences to introduce the topic. In this study, we looked at the first three sentences of each essay to code this lead-in.

Beyond the analysis of the organizational structure of the scripts, a corpus analysis was conducted to help us detect the differences and similarities in lexical items between the two groups. Using the AntConc software (Anthony, 2012), the "Top 100 lexical words" and "Top 100 functional words" were generated for both groups. The two lists were then compared and contrasted leaving aside those words that were directly related to the content of the two prompts, e.g., "daycare" and "urbanization."

4. Results

The results of both preliminary and further text analysis yielded indices of general writing performance for the two groups. Table 3 shows the output produced by the text analyzer and Table 4 shows the output of Coh-Metrix.

Table 3 reveals that the comparison group's average essay length, lexical density ($p < .05$) and Fog index[5] ($p < .05$) were greater than that of the experimental group. However, Table 4 indicates that the experimental group outperformed the comparison

[5] Fog Index estimates the number of years of education that readers need to understand a given text. It is measured by calculating the average number of words per sentence and the number of difficult words and generating an index from these measures.

Table 3
Results of automated text analysis (Indices of essay quality for experimental and comparison groups).

Group	Experimental (n = 65)	Comparison (n = 64)	Significance level *p*
Average length (no. of words)	228.35	242.11	
Lexical density	10.34%	11.54%	< .05
Fog index	8.6	9.8	< .05

Table 4
Results of Coh-Metrix analysis (Indices of essay quality for experimental and comparison groups).

Group	Experimental (n = 65)	Comparison (n = 64)	Significance level *p*
Syntactic complexity M (SD)	4.10(0.34)	3.64(0.31)	< .05
Lexical diversity M (SD)	75.69(3.98)	68.85(5.60)	< .05
Word frequency M (SD)	3.12(0.02)	3.07(0.03)	

Table 5
Descriptive statistics of metadiscourse use for both experimental and comparison groups.

Variable	Group			
	Experimental (N = 65)		Comparison (N = 64)	
	M	SD	M	SD
Textual metadiscourse				
Transitive markers	2.80	1.51	3.05	1.36
Frame markers	2.30	1.84	1.70	1.45
Code glosses	0.30	0.47	0.30	0.47
Interpersonal metadiscourse				
Hedges	1.66	1.43	0.85	0.93
Attitude markers	1.60	0.75	1.25	0.79

group in syntactic complexity ($p < .05$) and lexical diversity ($p < .05$); but in terms of word frequency, the comparison group demonstrated more proficiency because they used words that occur less frequently. The six indices in Tables 3 and 4 are all meant to reflect the writing quality of the essays. Note that it is possible that an essay has high lexical diversity, but low lexical density (Johansson, 2008); however, the ostensible contradictions between the text analyzer outputs and Coh-Metrix outputs will be discussed further in the next section.

Table 5 shows that the experimental group used significantly more hedges than the comparison group in their essays ($p < .05$; with an effect size Cohen's $d = 0.67$). Hedges that were used in the experimental group essays include: *may, might, maybe, perhaps, seem to, to some degree* and *personally*. Hedges used by the comparison group included: *maybe, to some extent* and *in a way*. The experimental group also used significantly more frame markers ($p < .05$; $d = 0.36$) and attitude markers ($p < .05$; $d = 0.45$). However, the comparison group used more transitive markers ($p < .05$; $d = 0.27$). The means of code glosses used by the two groups were the same, and these were the least frequently used among the five subcategories of metadiscourse by both groups.

Notably, there was a marked contrast in the way participants began their essays. It was found that in 75% (49 out of 65) of the experimental group essays, the topic was introduced without any preamble, e.g., "In my opinion, I think there are three reasons about this phenomenon;" in contrast, in the comparison group, 52 out of 64 essays (81%) began in a less direct, and arguably, more academically conventional way by hooking the reader with the provision of some background information, e.g.,

> Recently, there is a heated discussion about whether urbanization makes our life better. As we all know, China is developing very fast. As a consequence, the rate of urbanization is increasing. This changes our lives greatly. In my opinion, urbanization makes our life better in many aspects.

As for the organizational structure of the scripts, in the comparison group, 98% of the essays (n = 64) followed a three-part argumentation model, namely, an introduction-body-conclusion model. In almost all cases, the writer's stance was presented in the introduction and this stance was supported by evidence in the body. In the conclusion, the stance was restated to a greater or lesser extent. For the experimental group, however, the three-part structure was adopted in only 15% (n = 10) of the essays. Sixty percent (n = 39) included two or three paragraphs with no introduction or conclusion, and the remaining 25% (n = 16) either lacked an introduction or conclusion.

Table 6
Selected lexical differences between experimental and comparison groups.

Word type	Experimental Group			Comparison Group	
	Word	Rank	Frequency	Rank	Frequency
Lexical	think	10	115	12	67
	if	25	88	49	33
Functional	should	27	82	36	47
	first	40	41	58	23

Remarkable differences and similarities in lexical items were also noted by comparing the "Top 100 lexical words" and "Top 100 functional words" of the essays written by the two groups. Tables 6 and 7 respectively show some of the more notable differences and similarities in the two groups. "Think" was used 115 times in the experimental group compared to 67 in the comparison group. As to the context in which the word was used, the concordancer revealed there was also significant variation. Seventy-seven percent of the comparison group's usage of "think" was to express personal opinions or ideas, e.g., "I don't think urbanization is making our life better." For the experimental group, however, only 45% of cases were used in this way, while 55% of the instances of "think" were used in an attributive sense, i.e., ascribing other people's thoughts or beliefs, e.g., "this would lead to a misunderstanding of parents who think that money will solve everything." Thus, the word "think" was used by a significantly higher percentage of students in the experimental group ($p < .05$) to project thoughts onto others.

Perhaps the most noteworthy frequency difference was regarding the usage of the word "if," used in a hypothetical sense. For the experimental group, "if" was the 25th most common functional word used 88 times, while for the comparison group, "if" was used only 33 times, ranking 49th. A rereading of the contexts of sentences where the word "if" appeared in the scripts of the experimental group revealed that in almost all cases (n = 87), the writer was presenting hypothetical arguments, e.g., "if I have money, I would be late, [sic] and I would not obey the discipline," and "if they can pay for being late, they will not feel ashamed of themselves." In these two examples among many others in the experimental scripts, the authors not only conjectured parents' psychology, but also predicted potential consequences. In the comparison group, the frequency of hypothetical arguments with the word "if" (31 cases) or in other forms (two cases of hypothesis with modals) was much lower. For instance, a student wrote, "how nice if we keep living in the countryside!" and another wrote, "if we really want urbanization to make our life better, there is still a long way to go." By examining all the cases of the use of "if," we saw that the comparison group also made hypothetical arguments, but at a lower frequency with less variety. Specifically, the comparison group did not project thoughts onto others as the experimental group did.

5. Discussion

5.1. Interpretations of the seemingly contradictory findings

The findings in this study can be interpreted in two ways. Using standardized indices, the comparison group outperformed the experimental group in aspects including essay length, lexical density, Fog index and word frequency. In other words, the comparison group collectively wrote longer essays that were more conventionally academic in nature based on various standard measures. This may have been due to the students' familiarity with the type of prompt as well as the object of inquiry (socio-cultural issue) in addition to having frequent practice with evaluative and hortative rhetorical functions. Studies have shown that topic familiarity and prompt type exert significant positive effects on the writing performance of students across proficiency levels (Stapleton, 2001; He & Shi, 2012). Similarly, the more sophisticated lead-ins in the scripts of the comparison group again may have been products of repeated preparation. Another possible explanation for the longer and more academically conventional essays written by the comparison group is based on cognitive writing research, which notes that a reduced cognitive load (familiarity and preparation effects) can enable students to translate their thoughts with more ease and confidence, and hence write in a less stressful way (Fayol, 2012; Flower & Hayes, 1981). In our case, the familiar topic more or less freed writers from attending to their syntax, which meant a reduced burden on working memory for the comparison group. In essence, lower level thinking processes may lead to smoother writing, while more complex thinking causes surface disruptions, but more interesting content.

However, the experimental group essays demonstrated a higher level in terms of lexical diversity (use of different word types) and

Table 7
Selected lexical similarities between experimental and comparison groups.

Word type	Experimental Group			Comparison Group	
	Word	Rank	Frequency	Rank	Frequency
Lexical	feel	67	16	84	14
Functional	can	8	211	11	201
	because	39	44	42	42
	however	74	20	56	24

syntactic complexity (number of words before the main verb), which is suggestive of a higher cognitive demand on the part of the writer (Flower & Hayes, 1981; McNamara et al., 2010). Since the whole writing process involves intensive cognitive activity, for a writer to compose more complex words and sentences she needs more cognitive capacity (Flower & Hayes, 1981). The greater cognitive effort imposed by the unconventional writing prompt for the experimental group might have resulted in more diverse rhetorical and pragmatic dimensions in the students' essays (MacArthur & Graham, 2016), which could also be viewed as providing an opportunity for student respondents to exercise their latent critical thinking abilities.

Qualitative analysis of the scripts revealed that the comparison group used similar sentences in their essays with some writers copying sentences from the writing prompt. Many students in the comparison group, for example, began their essays with a similar, or even identical sentence. This may have caused the low scores in syntactic complexity and lexical diversity for the comparison group. The same may be said of the collective uniformity of organizational patterns noted in the paragraphs of the comparison group's scripts. In contrast, (the second interpretation) receiving an unfamiliar task and topic, which the exploratory prompt was meant to be, may have disadvantaged the writing performance of the students in the experimental group. Because of the shorter length, lower scores on the two standard measures, and inconsistent organization, the experimental group was clearly outperformed. Nevertheless, breaking the constraints of the "three-part" structure may be considered another positive cognitive payoff for the experimental group. We speculate that the diversity in the structure of essays could also be a sign of and stimulus for critical thinking.

However, this does not mean that the students in the experimental group were hindered in their thinking and arguing abilities. The higher frequency of hedges by the experimental group could suggest that the problem-solving prompt is more stimulating in terms of critical thinking. Although hedges serve to reduce an author's commitment to a statement, they may reflect the author's speculation, imagining other people's minds and interpreting other people's behavior, which typifies high-order thinking. The use of hedges has been increasingly studied in undergraduate academic writing (e.g., Aull, 2015; Aull & Lancaster, 2014). Researchers suggest that using hedges may be considered a crafted way of balancing conviction and caution by relatively advanced writers (Aull, 2015; Hyland, 2000), but the nature of hedge use is still inconclusive.

Another illustration of possible higher-order thinking was the greater use of the words "if" and "think" by the experimental group. "If" is often used to make a prediction or generate a hypothetical situation. Take-a-stand-oriented prompts may be less likely to stimulate this type of thinking or their associated linguistic structures. Likewise, the marked difference in the way the two groups used the word "think" is revealing. Rather than simply using this word to express personal beliefs as the comparison group tended to do, a significantly higher percentage of writers in the experimental group projected thoughts onto others - certainly a more challenging mental exercise both logically and linguistically. Undeniably, the isolation of two words where differences appeared between the two groups is not sufficient to make any generalizations beyond the present study; however, in the exploratory spirit, these examples illustrate how areas of omission, both logical and linguistic, can emerge when prompts gravitate towards the one-size-fits-all approach apparently taken by the high-stakes tests, especially TOEFL and IELTS.

5.2. The role of writing prompts in cultivating students' argumentative writing and critical thinking abilities

Due to the considerable washback of high-stakes test writing prompts, the narrowness of rhetorical function and object of enquiry can be observed in writing pedagogy (Ketter & Pool, 2001; Moore & Morton, 2005). In the composition class at the university where the experiment took place, the object of enquiry of a typical prompt is usually sociocultural, and the rhetorical function always requires students to take a stand on an issue and support it, i.e., evaluative and/or hortative; to our knowledge, this might be the case in most writing classes at Chinese universities. Consequently, students have inadequate practice in alternative objects of enquiry and rhetorical function, which may partly explain why some Chinese students studying overseas struggle to complete university writing assignments even though they score high in standardized writing tests.

In one study by Wang (2010), Chinese university students were invited to select prompts that they would like to respond to from a collection of topics. Wang (2010) found that although students' selection differentiates to some degree from actual prompts used in high-stakes test prompts, both students' prompt choice and test prompts were confined to a narrow scope of society and life experience. This narrow scope may again reflect test washback on classroom instruction as well as assessment practice in mainland China. Despite the negative effect outlined here, the flipside is the positive potential that the writing prompt has if new objects of enquiry and rhetorical functions are considered for the purposes of enhancing students' analysis, problem-solving and other critical thinking abilities.

In this exploratory study, a writing prompt with problem-solving as the main rhetorical function, and behavioral economics as the object of enquiry elicited more diverse language and critical thinking skills that were already presented in the student writers. This does not imply that the comparison group, by responding to the conventional prompt, did not display critical thinking skills in their writing. The comparison group essays did include elements of critical thinking such as evaluation, recommendation, and explanation to various extents. However, the experimental group involved a wider spectrum of critical thinking skills such as interpretation, prediction, and problem-solving. Actually, the skill of problem-solving is considered an essential element of both the cognitive and dispositional dimensions of critical thinking (e.g., Facione, 2015; Halpern, 1998, 2003). The thinking processes involved in problem-solving entail reflective and procedural mental activities, which demand more cognitive competence and higher-order thinking, and hence need practicing and cultivating (Halpern, 2003; Perkins & Tishman, 2001). Thus, we argue that such problem-solution and behavior-interpretation in the writing tasks has important educational implications for critical thinking development.

5.3. Limitations

An evident limitation of this study concerns the writing prompts generated for the two groups, which compared responses to conventional and experimental prompts. Clearly, the two sets of scripts resulting from these prompts are distinct, and any patterns noted between the two groups are unique, and as is customary in exploratory research, can be taken as indicative only. Indeed, because the experimental prompt did not specify a genre, there was considerably more randomness in the structure of responses from students, which, if this new text type were to be adopted, would require associated genre instruction.

Another issue concerns the length of the essays supplied by the students. It could be argued that the short essays collected in the present study were not long enough for the type of analysis performed, or for the students to develop proper arguments. On the other hand, such limitations are common when dealing with the timed, written responses to a previously unseen prompt in a test situation.

Our experimental prompt, which explored responses from students, also largely ignored the practical issue of scoring reliability. Thus, it could be argued that straying into an alternative object of enquiry, i.e., behavioral economics, approaches territory beyond the realm of language, although similar arguments could be made about those presently used, i.e., those from the socio-cultural domain. And this point regarding practical issues also raises concerns about scoring in high stakes tests: if such tests were to adopt a greater number of rhetorical functions and objects of inquiry, it would inevitably create reliability challenges due to the difficulty of creating scoring rubrics that encompass a wider range of texts. Or alternatively, if the specific type of prompt advocated in this study, i.e., one that encourages critical thinking, were adopted, then test takers may well begin preparing specifically for them, eventually resulting in more cookie-cutter conventional responses.

Two final limitations concern our framing of the present study as "exploratory." An argument could be made that we initiated the study with preconceived notions that writing prompts can influence students' responses, leading to post hoc reflections in this Discussion. This argument does have some merit as we did expect some differences to arise. However, the nature of these differences were unknown to us, and did not exist in the literature. Thus, the experimental prompt was meant to explore possibilities without intending to provide any conclusive answers. Secondly, ideally, after the first iteration of the study, the two groups would have reversed their roles and produced a second set of scripts for analysis, perhaps providing a more reliable measure of writing quality. Thus, in both of these senses, we framed the study as exploratory.

6. Conclusion

Condon (2009), p. 143 observes that one of the most challenging tasks in writing assessment is devising a workable writing prompt. "Factors of class, region, nationality, language, and culture stand in the way of constructing such prompts." In response to this challenge, he proposed a type of prompt that calls on students to infuse their personal experience. In the same vein, the present study also seeks to unsettle the status quo. Because of concerns that the writing prompts in high-stakes language exams tend to converge around a narrow set of rhetorical functions and objects of inquiry, the present study sought to compare responses from students to a conventional prompt and an exploratory one. An analysis of the scripts from a comparison group who responded to a conventional prompt revealed that although this group performed better in terms of certain linguistic indexes, the test-takers' organization and language use was restricted to certain patterns and usages when compared to an experimental group that responded to the exploratory prompt. Given the impact that writing prompts in high-stakes exams have on language curriculums and classroom instruction pedagogy, these findings suggest a need to widen the scope of genres and topics in writing prompts that students are expected to respond to. And while this exploratory study investigating the impact of one rhetorical function and one object of enquiry on student writing can only hint at alternative types of patterns of writing and thinking, it gives impetus for research into broadening the scope of prompts in high stakes tests. New research may indicate there is potential to expand beyond the present narrow form of writing preferred by high stakes tests into new text types and topics which could trigger diversity and enhanced thinking processes.

References

Alderson, J. C., & Wall, D. (1993). Does washback exist/. *Applied Linguistics, 14*(2), 115–129.
Andrews, S. (2004). Washback and curriculum innovation. In Cheng L, & Watanabe Y (Eds.). *Washback in language testing: Research contexts and methods* (pp. 37–50). Mahwah, NJ: Lawrence Erlbaum.
Anthony, L. (2012). *AntConc (ver. 3.3.5) [software].* Retrieved on March 19, 2018 fromhttp://www.antlab.sci.waseda.ac.jp/.
Ariely, D. (2008). *Predictably irrational: The hidden forces that shape our decisions.* New York: HarperCollins.
Aull, L. L. (2015). Connecting writing and language in assessment: Examining style, tone, and argument in the U.S. Common core standards and in exemplary student writing. *Assessing Writing, 24*, 59–73.
Aull, L. L., & Lancaster, C. I. Z. (2014). Linguistic markers of stance in early and advanced academic writing: A corpus-based comparison. *Written Communication, 31*, 151–183.
Beck, S. W., & Jeffery, J. V. (2007). Genres of high-stakes writing assessments and the construct of writing competence. *Assessing Writing, 12*, 60–79.
Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly, 45*, 5–35.
Carroll, D. W. (2007). Patterns of student writing in a critical thinking course: A quantitative analysis. *Assessing Writing, 12*, 213–227.
Carroll, R. T. (2004). *Becoming a critical thinker: A guide for the new millennium.* London: Pearson Custom Publishing.
Cheng, L. (1998). Impact of a public English examination change on students'perceptions and attitudes toward their English learning. *Studies in Educational Evaluation, 24*, 279–301.
Condon, W. (2009). Looking beyond judging and ranking: Writing assessment as a generative practice. *Assessing Writing, 14*, 141–156.
Condon, W., & Kelly-Riley, D. (2004). Assessing and teaching what we value: The relationship between college-level writing and critical thinking abilities. *Assessing Writing, 9*, 56–75.
Crismore, A., Markkanen, R., & Steffensen, M. S. (1993). Metadiscourse in persuasive writing: A study of texts written by American and Finnish university students.

Written Communication, 10, 39–71.
Crusan, D. (2010). *Assessment in the second language writing classroom.* Ann Arbor, MI: University of Michigan Press.
Durst, R. K. (1987). Cognitive and linguistic demands of analytical writing. *Research in the Teaching of English, 21*(4), 347–376.
Eckstein, M. A., & Noah, H. J. (1993). *Secondary school examinations: International perspectives on policies and practice.* New Haven, CT: Yale University Press.
Education Bureau (2014). *Resource Pack for Economics Curriculum (Secondary 4–6).* Retrieved on Sept. 13, 2018, from https://www.edb.gov.hk/attachment/en/curriculum-development/kla/pshe/references-and resources/economics/Econ_Essay_Booklet_(14_Nov_6).pdf.
Facione, P. A. (2015). *Critical thinking: What it is and why it counts. Retrieved on March 19, 2018.* fromwww.insightassessment.com/Resources/Tools-For-Teaching-For-and-About-Thinking/Critical-Thinking-What-It-Is-and-Why-It-Counts/Critical-Thinking-What-It-Is-and-Why-It-Counts-PDF.
Fayol, M. (2012). Cognitive processes of children and adults in translating thought into written language in real time. In V. W. Berninger (Ed.). *Past, present, and future contribution of cognitive writing research to cognitive psychology* (pp. 27–60). NY: Taylor & Francis.
Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication, 32*(4), 365–387.
Furtak, E. M., & Ruiz-Primo, M. A. (2008). Making students' thinking explicit in writing and discussion: An analysis of formative assessment prompts. *Science Education, 92*, 799–824.
Gneezy, U., & Rustichini, A. (2000). A fine is a price. *Journal of Legal Studies, 29*, 1–17.
Green, A. (2006). Watching for washback: Observing the influence of the international English language testing system academic writing test in the classroom. *Language Assessment Quarterly, 3*, 333–368.
Halpern, D. F. (1998). Teaching for critical thinking: Helping college students develop the skills and dispositions of a teaching critical thinking for transfer across domains: Dispositions, skills, structure training, and metacognitive monitoring. *American Psychologist, 53*(5), 449–455.
Halpern, D. F. (2003). *Thought and knowledge: An introduction to critical thinking* (4th ed). Mahwah, NJ: Lawrence Erlbaum.
He, L. (2010). The graduate school entrance English examination. In L. Cheng, & A. Curtis (Eds.). *English language assessment and the Chinese learner* (pp. 145–157). London: Routledge.
He, L., & Shi, L. (2012). Topical knowledge in ESL writing. *Language Testing, 29*, 443–464.
Hinkel, E. (1995). The use of modal verbs as a reflection of cultural values. *TESOL Quarterly, 29*(2), 325–343.
Hinkel, E. (2002). *Second language writers' text: Linguistic and rhetorical features.* Mahwah, NJ: Lawrence Erlbaum.
Hinkel, E. (2009). The effects of essay topics on modal verb uses in L1 and L2 academic writing. *Journal of Pragmatics, 41*(4), 667–683.
Hyland, K. (2000). Hedges, boosters and lexical invisibility: Noticing modifers in academic texts. *Language Awareness, 9*(4), 179–197.
Hyland, K. (2005). *Metadiscourse: Exploring interaction in writing.* London: Continuum.
Johansson, V. (2008). Lexical diversity and lexical density in speech and writing: A developmental perspective. *Lund Working Papers for Linguistics, 53*, 61–79.
Ketter, J., & Pool, J. (2001). Exploring the impact of a high-stakes direct writing assessment in two high-school classrooms. *Research in the Teaching of English, 35*, 344–393.
Lee, H., & Anderson, C. (2007). Validity and topic generality of a writing performance test. *Language Testing, 24*(3), 307–330.
Liu, F., & Stapleton, P. (2015). Writing prompt convergence in high-stakes tests: Exploring rhetorical functions and objects of enquiry. *Studies in Educational Evaluation, 47*, 68–75.
Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly, 45*, 36–42.
MacArthur, A. C., & Graham, S. (2016). Writing research from a cognitive perspective. In A. C. MacArthur, S. Graham, & J. Fitzgerald (Eds.). *The handbook of writing research* (pp. 24–40). (2nd ed). NY, London: The Guilford Press.
McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic features of writing quality. *Written Communication, 27*, 57–86.
Moore, T., & Morton, J. (2005). Dimensions of difference: A comparison of university writing and IELTS writing. *Journal of English for Academic Purposes, 4*, 43–66.
Palmer, W. (2012). *Discovering arguments: An introduction to critical thinking, writing, and style* (4th ed). Boston, MA; Hong Kong: Prentice Hall.
Perkins, D. N., & Tishman, S. (2001). Dispositional aspects of intelligence. In J. M. Collis, & S. Messick (Eds.). *Intelligence and personality: Bridging the gap in theory and measurement* (pp. 233–257). Mahwah, NJ: Lawrence Erlbaum.
Popham, W. J. (2001). Teaching to the test. *Educational Leadership, 58*(6), 16–20.
Popham, J. (1987). *The merits of measurement-driven instruction.* Phi Delta Kappa May679–682.
Qi, L. (2007). Is testing an efficient agent for pedagogical change? Examining the intended washback of the writing task in a high-stakes English test in China. *Assessment in Education: Principles, Policy & Practice, 14*(1), 51–74.
Ramage, J. D., Bean, J. C., & Johnson, J. (2010). *Writing arguments: A rhetoric with readings* (8th ed). New York, NY; Hong Kong: Pearson Longman.
Reid, J. (1990). Responding to different topic types: A quantitative analysis from a contrastive rhetoric perspective. In B. Kroll (Ed.). *Second language writing: Research insights for the classroom* (pp. 191–209). Cambridge: Cambridge University Press.
Shirkhani, S., & Fahim, M. (2011). Enhancing critical thinking in foreign language learners. *Procedia – Social and Behavioral Sciences, 29*(1), 111–115.
Shohamy, E. (2007). Language tests as language policy tools. *Assessment in Education: Principles, Policy & Practice, 14*(1), 117–130.
Spaan, M. (1993). The effect of prompt in essay examinations. In D. Douglas, & C. Chapelle (Eds.). *A new decade of language testing research* (pp. 98–122). Alexandria, VA: TESOL.
Stapleton, P. (2001). Assessing critical thinking in the writing of Japanese university students: Insights about assumptions and content familiarity. *Written communication, 18*, 506–548.
Tedick, D. J. (1990). ESL writing assessment: Subject-matter knowledge and its impact on performance. *English for Specific Purposes, 9*, 123–143.
Wang, D. (2010). Chinese students' choice of writing topics: A comparison between their self-selected topics and the writing prompts in large-scale tests. *The Journal of Asia TEFL, 7*(3), 165–187.
Weigle, S. C. (2002). *Assessing writing.* Cambridge: Cambridge University Press.
Yu, G. (2010). Lexical diversity in writing and speaking task performances. *Applied Linguistics, 31*, 236–259.

Fulan Liu (PhD) is a professor at Jiangxi Normal University, China. She has obtained her PhD degree from the Education University of Hong Kong, and her master's degree from the University of Melbourne. Her research interest is second language acquisition and academic writing.

Paul Stapleton is an associate professor at the Education University of Hong Kong.