

Secondary school students' discourse synthesis performance on Chinese (L1) and English (L2) integrated writing assessments

Xinhua Zhu¹

Guan Ying Li²

Choo Mui Cheong³

Guoxing Yu⁴

Xian Liao⁵

^{1,2}The Hong Kong Polytechnic University, Hong Kong SAR, People's Republic of China; ³The University of Hong Kong, Hong Kong SAR, People's Republic of China, ⁴University of Bristol, Bristol, United Kingdom; ⁵The Education University of Hong Kong, Tai Po N.T, Hong Kong

Abstract

This study investigates the relationship between three discourse synthesis skills (i.e., quotation, summarization, and connection) and students' overall integrated writing performance in Chinese, students' first language, and English, their second language. Hierarchical regression analyses revealed that 63.6% of the variance in students' overall Chinese integrated writing performance was accounted for by the three discourse synthesis skills, with connection and summarization contributed almost equally to the overall scores. In the English test, the three skills explained 47.9% of the variance. Cross-linguistic facilitation of the L1 discourse synthesis skills to the overall L2 integrated writing performance was observed, although the predictive strength of the three skills was comparatively low. Eye-tracking data together with subsequent stimulated-recall interviews illuminated the differences in students' approaches to discourse synthesis. Findings of the study support the decisive role of discourse synthesis abilities in both L1 and L2 integrated writing assessments. Implications for writing instruction are discussed.

Keywords Cross-linguistic facilitation · Discourse synthesis skills · Integrated writing assessment

Integrating writing has been recognized as one of the core competencies in academic discourse communities, in which writing skills are usually deployed in concert with other skills. Assessments integrating various language modalities have also been an integral part of language proficiency tests such as Test of English as a Foreign Language (TOEFL) and Canadian Academic English Language (CAEL), through which universities can ensure that students meet literacy requirements for higher education. Integrated writing assessment, by its very nature, is aimed at eliciting students' abilities to draw together information from multiple source materials to create new texts. There is a growing consensus that the ability to write from sources is not merely a summation of writing and other language skills; it requires "a reciprocal interaction" between language skills (Asención Delaney, 2008), representing a view of integrated writing assessment that is "holistic rather componential" (Plakans & Gebril, 2012, p. 18).

In Hong Kong, Chinese integrated writing assessment (i.e., a listening-reading-writing task) was introduced in 2005 for the first examination in 2007 (Hong Kong Examinations and Assessment Authority, 2005). Similar changes were also embodied in the new English language frameworks (Curriculum Development Council & Hong Kong Examinations and Assessment Authority, 2007a). The implementation of integrated writing assessment marked a significant change to language education in Hong Kong and echoed the city's language policy. The long-standing "biliteracy and trilingualism" policy aims to nurture learners who are proficient in speaking Cantonese, English, and Putonghua and can read and write fluently in Chinese and English (Education Bureau, 2010). It was stated clearly in the subsequent curriculum guidelines that adopting an integrated approach to language learning in both Chinese and English is at the core of language education at the senior secondary school level to develop language competence holistically, and more importantly, to provide the foundations for higher education (Curriculum Development Council & Hong Kong Examinations and Assessment Authority, 2007a, 2007b). Although integrated writing stands as a key competency in higher education and as a prominent part of Hong Kong's bilingual education system, little has been done in the local context to examine students' writing performance since the first official examination in 2007. Researchers have stressed the importance of comparing L1 and L2 integrated writing performance in order to 'disentangle the differential effects of language proficiency and the ability to write from sources' (Cumming, Lai, & Cho, 2016, p. 54). To this end, this study set out to investigate students' ability to use source texts in both Chinese (students' L1) and English (their L2) integrated writing assessments. Previous studies have also pointed out the cognitively demanding nature of integrated writing (Asención Delaney, 2008; Plakans & Gebril, 2012). Apart from writing assessments, the current study also adopts eye tracking technology in concert with stimulated recall interviews to delve into the processes and factors that would affect students' writing performance.

Literature review

Source use in integrated writing assessment

Integrated writing assessments require cognitive operations that are distinct from assessments of a single skill (Asención Delaney, 2008; Grabe, 2001). The com-

posing process of integrated writing involves synergy between language modalities as well as between information from multiple sources. Zhu, Li, Yu, Cheong, and Liao (2016) found that two factors in an independent listening task, evaluation and creation, working in concert, significantly predicted students' performance on a listening-reading-writing task. Cheong, Zhu, and Liao (2018) further indicated that three higher-order thinking skills—elaborating, evaluating, and creating in both the reading and listening task, were significantly correlated with integrated writing performance. These studies confirmed the interdependence of writing and other language skills in integrated writing tasks. Spivey and King (1989) referred to this complex meaning-making process as “discourse synthesis”. Discourse synthesis requires writers not only to comprehend source information but to transform the given sources by employing three fundamental operations of organizing, selecting, and connecting to create their own written products.

Inspired by Spivey and King's (1989) conceptualization of discourse synthesis, writers' use of source materials has gained increasing attention in the field of language assessment. The study by Sawaki, Quinlan and Lee (2013) revealed that writers' ability to comprehend the reading and listening materials and their own written products was a significant factor accounting for integrated writing performance. Prior research has also tapped into verbatim source use and integration styles (e.g., quoting, paraphrasing, or summarizing). In a study of discourse synthesis processes of second language writers, Plakans (2009) found that higher frequencies of discourse synthesis processes led to greater proportions of source use in writing.

Yang and Plakans (2012) further indicated that the use of discourse synthesis strategies had direct positive effects on integrated writing performance. Responding to integrated writing assessments with copying or patchwriting strategies, on the other hand, had negative impact on test performance. Plakans and Gebril (2012) employed a mixed-methods approach to investigate source text use in an L2 integrated writing assessment. Various patterns of source use emerged from the think aloud protocols and post-writing questionnaire, showing that L2 writers used source materials not only to gain ideas about writing tasks but also for language support. Moreover, constant interaction between the source texts and the participants' first language was observed; the participants tended to rely on their first language to comprehend source materials and formulate ideas in their composing processes.

Language proficiency was also found to be a critical factor resulting in differences in discourse synthesis performance; lower language proficiency seemed to inhibit the employment of discourse synthesis skills (Plakans, 2009). In a TOEFL integrated writing task, Plakans and Gebril (2017) found that low-performers seemed to focus more on the reading passage and pay less attention to the listening, whereas high-performers exhibited more balanced summarization of sources. Interestingly, in an examination of summarization strategies employed by L1 and L2 writers, Keck (2014) observed similarities between the two groups. Both L1 and L2 writers tended to present information in the same order that the selected excerpts appeared in the original text. Moreover, L2 writers used paraphrasing strategies almost as frequently as their L1 peers did. In this study, however, the quality of summaries composed by the students was not considered. It was unclear whether quantitatively similar summarization practices could contribute to similar writing quality.

Even though it is well established that discourse synthesis skills can characterize

integrated writing performance at different ability levels, it remains uncertain to what extent discourse synthesis skills influence writing quality and the effects of individual discourse synthesis skills on integrated writing scores.

Relationships between L1 and L2 writing

Cummins' (1979) hypotheses of linguistic interdependence and linguistic threshold have led to a burgeoning line of inquiry into the relationship between L1 and L2 literacy skills, showing positive effects of L1 on L2 (Kim & Piper, 2019; Shum, Ho, Siegel, & Au, 2016; Sparks, Patton, & Luebbbers, 2019). It is commonly asserted that skilled L2 writers tend to be skilled in L1 writing and vice versa (Deygers, Van den Branden, & Peters, 2017; Leki, Cumming, & Silva, 2008; Whalen & Ménard, 1995). Comparison studies of the processes of L1 and L2 writing have illuminated the differences between more proficient writers and their counterparts. Apart from linguistic knowledge, writer's level of cognitive development also played a prominent role in written discourse production (Whalen & Ménard, 1995).

Research on biliteracy acquisition has lent support to the possible transfer of L1 reading-related skills to L2 (Keung & Ho, 2009; Wang, Perfetti, & Liu, 2005). Comparing students' writing in their first language (Dutch) and second language (English), Schoonen et al. (2003) indicated that scores on L1 writing was highly correlated with the scores on L2 writing, and that L2 writing proficiency was subjective to linguistic knowledge and lexical retrieval speed, whereas in L1 writing, the impact of metacognitive knowledge was more evident. The level of attainment in L1 literacy skills was predictive of the development of L2 (Sparks, 2012). L1 reading achievement had significant contribution to L2 reading comprehension and overall L2 proficiency (Sparks, Patton, Ganschow, & Humbach, 2012). Similar findings were observed in a longitudinal study in which L1 proficiency in elementary school accounted for the differences in L2 literacy skills in high school (Sparks, Patton, Ganschow, & Humbach, 2009).

Cross-linguistic facilitation from L1 to L2 was also reported in the study by Savage, Kozakewich, Genesee, Erdos and Haigh (2017), showing that Grade 1 language comprehension scores were a strong predictor of writing accuracy in Grade 6 both intra-linguistically and cross-linguistically, whereas the predictive strength of the language comprehension variables on the quality of persuasive writing was comparatively weak. In brief, previous studies have provided consistent evidence for the cross-linguistic transfer of literacy skills among bilingually educated students. The level of attainment in L1 will impact students' L2 proficiency in both receptive and productive skills (Sparks, 2012). Nevertheless, it is worth noting that in previous studies, the effects of L1 literacy skills on L2 writing performance were either not significant (Sparks et al., 2012) or relatively weak (Savage et al., 2017). Moreover, most of the studies focused primarily on reading-related skills (e.g., word decoding, spelling, vocabulary). Rarely did they explore the cross-linguistic relationship of integrated writing skills. Further exploration is needed to extend current research on cross-linguistic facilitation.

Eye movement and cognitive processes in language assessment

When writing from sources, writers are also readers. They are engaged in the discourse settings in which the source materials and their writing interact reciprocally

and constantly. This complex meaning-making process imposes a high cognitive load on writers to process information both intra-textually and inter-textually (Grabe, 2001; Segev-Miller, 2007). Recently eye tracking technology has been applied to uncover the cognitive processes elicited by language assessments, resting mainly on the eye-mind hypothesis that “there is no appreciable lag between what is being fixated and what is being processed” (Just & Carpenter, 1980, p. 331). Researchers can infer from gaze duration the allocation of attention in cognitive activities, such as reading and writing. In their study of the cognitive processes elicited by reading tests, Bax and Chan (2016) observed that successful candidates were more strategically capable in terms of identifying and focusing on key areas of each test item and text than their counterparts.

Similar findings were found in Yu, He, and Isaacs (2017). The researchers examined students’ cognitive processes in graph-based writing tasks in International English Language Testing System (IELTS). High performers visited more frequently and fixated longer on the key information and spent less time on the task instruction than low performers did. Nevertheless, as the researchers noted, the relationships between participants’ eye-movements and their writing ability were inconclusive (Yu et al., 2017). No consistent pattern was observed in the study, thereby requiring further research to examine the relationships. Given the complexity of integrated writing assessment, researchers need to examine not only the product but also the process of integrated writing. Eye tracking technology allows researchers to observe writing processes in a relatively naturalistic setting without burdening participants with extra cognitive load caused by other methods such as concurrent think-aloud. Observations of students’ source use would provide illuminating insights that help account for differences in integrated writing performance as suggested in the previous studies. In addition, the eye movement videos recorded throughout the processes are useful in facilitating follow-up interviews with test-takers. It is expected that eye movement data together with interviews would deepen researchers’ understanding of factors that affect integrated writing performance.

The present study

In light of the issues arising from previous research and the integral role of discourse synthesis in the local assessment regime, the present study focused on the effects of three discourse synthesis skills (i.e., quotation, summarization, and connection) on integrated writing performance. The three skills were defined as follows. Quotation refers to verbatim use of source materials, which is a key indicator used to differentiate students’ source use performance in prior research (Yang & Plakans, 2012). Additionally, it has also been indicated in the local examination reports that low performing students tended to cite source information with excessive direct copying or patchwriting (Hong Kong Examinations and Assessment Authority, 2010, 2012). Given that the repercussion of inappropriate use of quotation would be plagiarism which could risk students’ academic integrity, we chose quotation as one of the indicators to differentiate discourse synthesis performance.

Summarization is concerned with writers' ability to condense source information into a brief statement that "reflects the gist of the discourse" (Hidi & Anderson, 1986, p. 473). It has been one of the major criteria in the local assessment framework and established as a valid construct that illuminates the differences in organizational quality in discourse synthesis (Plakans & Gebril, 2017; Yu, 2009, 2013). Connection, termed following Spivey and King's (1989) framework, taps into the core competency in integrated writing—the ability to integrate information selected from source materials into coherent unities (Cumming et al. 2016; Plakans & Gebril, 2017). In avoiding overlap between summarization and connection, in the present study, summarization was operationalized in a single text context, in which students were required to condense the main idea of a single source material. Connection was focused on intertextual processing in which students were required to synthesize key information from across multiple sources to demonstrate their understanding of differing perspectives on a given topic.

The investigation of the relationship between discourse synthesis skills and students' integrated writing performance in L1 (Chinese) and L2 (English) was guided by the following two research questions:

1. What are the effects of discourse synthesis skills on students' integrated writing performance? Do the effects differ across L1 and L2 assessments?
2. What is the relationship between students' L1 and L2 discourse synthesis performance on integrated writing assessments? Is there a cross-linguistic effect of L1 discourse synthesis skills on L2 writing?

Method

The present study investigated students' integrated writing performance in two test settings: paper-based and computer-based tests. The paper-based tests allowed us to examine the effects of discourse synthesis skills on integrated writing performance with a larger sample size. The computer-based tests, on the other hand, enabled a closer investigation of the relationship between the writing performance and the composing processes of individual participants. The two test settings are described in detail below.

Test 1: paper-based assessments

Participants

The present study involved 145 Secondary 4 students from two secondary schools in Hong Kong. Here Secondary 4 students referred to 10th graders who are in their fourth year in high school. The participants are native Chinese learners, and many of them have been learning English as a second language formally since first grade. We chose Secondary 4 students for the present study because integrated writing is usually introduced at this phase. At the time when the present study was carried out,

students had learned about the requirements of integrated writing and were familiar with the assessment format. The gender distribution was 53 females and 92 males with an average age of 15.89 ($SD = 0.77$). The two schools, including one Band 2 school and one Band 3 school, formed a sample that might reflect the academic achievement levels of the majority of students in the local context (school banding is a system commonly used in Hong Kong to represent students' overall academic achievement, with Band 1 being the best, followed by Band 2 and Band 3).

Instruments

A set of integrated writing tests and integrated writing scoring rubrics were used in the present study to measure and evaluate students' discourse synthesis skills as well as the overall integrated writing performance in both L1 and L2 settings.

The integrated writing tests

A Chinese integrated writing test (CIW) and an English integrated writing test (EIW) were developed for the present study. These tests were paper-based and formatted in alignment with the current HKDSE Chinese Language Paper 3 (i.e., a listening-reading-writing task). We followed the format of HKDSE because tasks integrating listening, reading, and writing modalities are reflective of authentic language environment, and tasks in such or similar format have been applied in international language assessments, such as TOEFL iBT. In each test, the test materials consisted of six reading passages, an audio recording, and a writing prompt. The first text was a brief description of task requirements, and the rest of the five texts were selected to present different perspectives on a topic of interest. These texts were either consistent with or contradictory to each other in terms of position. The selection criteria for the reading passages for the CIW and EIW were basically the same with slight differences in text type. Both tests included news articles narrating incidents related to a given topic, a bar chart describing a trend in a phenomenon of interest, a discursive essay arguing the upside and downside of an issue of interest, and proverbs/quotations. The audio recordings were in the form of discussion forum comprised of one host and two invited guests discussing a given topic. The content of the recording was relevant to, but not overlapping with the reading passages. The writing tests required students to write a speech to summarize the different perspectives presented in the readings and the recording and to express personal views on a given topic. Speech writing was chosen because it is a common text type in the local school context. Students at secondary school level are familiar with it. Local teachers and experts in language education and assessment were consulted regarding the clarity and appropriateness of the tests.

The topic of the Chinese test was about priority seats on public transportation, which is very familiar to most of the students living in Hong Kong. The six reading passages amounted to 1812 Chinese characters, encompassing: (1) a poster showing school anniversary celebration events and the requirements for the writing task, (2) a commentary arguing a new approach to priority seats on Hong Kong MTR, (3) a bar chart demonstrating the growth trend in the population aged 65 and above in

2011, 2015, and 2064, (4) two news articles with the first one illustrating a dispute over eligibility for priority seats in Taiwan, and the other reporting the results of a questionnaire survey about yielding seats in Hong Kong, (5) two ancient Chinese proverbs with translation, and (6) an illustration of ‘phubbers’ on the bus.

The topic of the English test was underage organ donation and transplantation based on a high-profile case that occurred in Hong Kong in 2017. The six texts amounted to 787 words, encompassing: (1) an email from a teacher indicating task requirements, (2) a bar chart showing the differing trends in living organ donation and deceased organ donation from 2012 to 2016, (3) an essay discussing the complexity of decision making involved in living organ donation, (4) a news article regarding underage liver donation, adapted from the aforementioned case happening in Hong Kong, (5) quotations from a discussion forum providing different views on the aforementioned 2017 incident, and (6) a news article reporting a case of medical malpractice, which led to a patient needing a liver transplant.

The integrated writing scoring rubrics

To evaluate students’ discourse synthesis skills and their overall integrated writing performance, we adapted the analytic integrated writing scoring rubrics (“Appendix”) from the marking rubrics used in Cheong et al. (2018). The adapted rubrics focused on the three discourse synthesis skills (i.e., quotation, summarization, and connection). With respect to students’ overall integrated writing performance, we used scoring rubrics adapted from the current HKDSE marking scheme (Hong Kong Examinations and Assessment Authority, 2017). The prevailing HKDSE scoring rubrics were built upon Zhu’s (2005) work for assessing integrated writing competence of secondary school students with regard to contextual awareness, source use, argumentation, and written expressions. The assessment framework has also been applied to assess students’ integrated writing performance in previous studies (Cheong, Zhu, Li, & Wen, 2019; Cheong et al., 2018; Zhu et al., 2016).

Procedures

The paper-based tests were carried out at the two local secondary schools. To complete the tests, students spent three minutes skimming the six reading passages prior to the listening part. Then they listened to a recording for approximately 12 min and wrote an essay in response to the writing prompt in 1 h. While the recording was only played once, the reading passages were available to students throughout the test.

With respect to the scoring procedures, each written text was marked by two raters who are experienced in teaching Chinese or English language with master’s degrees in the related fields. Prior to the actual marking, standardization meetings were held three times. Each lasted 2–3 h. The raters first met to have an intensive discussion on the scoring rubrics regarding the appropriateness and precision of the level descriptions. In the subsequent meetings, anchor essays that exemplified different proficiency levels in the rubrics were selected for trial marking and discussion. Moderations to the analytic scoring rubrics were made after the meetings. Each rater

then marked the entire set of written texts independently. The actual marking lasted about 7 days. On the second and the fourth day, interrater reliability was calculated for each marking group to examine the consistency of the judgements made by the raters. In the CIW, the inter-rater reliability estimates of the three discourse synthesis subscales ranged from 0.73 to 0.84, using Pearson product-moment correlation. The inter-rater reliability estimate of the overall score was 0.86. In the EIW, the inter-rater reliability estimates of the three discourse synthesis subscales ranged from 0.91 to 0.96. The inter-rater reliability estimate of the overall score was 0.84. A third rater reviewed the essays if the discrepancies in ratings occurred between the two raters (i.e., the scores assigned by the two raters differed by two points or above). The score assigned by the third rater was summed and averaged with the closest score assigned by the original rater.

Test 2: computer-based assessments

Participants

The participants of the computer-based tests were three male Secondary 4 students (i.e., 10th graders) from three secondary schools in Hong Kong. The three students were recruited on the recommendation of their Chinese language teachers to make sure that the students were familiar with word processing systems and at least one Chinese input method.

Computer-based integrated writing tests

Test materials mentioned previously in the section of paper-based tests, including the CIW, EIW, and the reading passages were delivered in web page format for the computer-based integrated tests. To ensure equivalence, the writing tasks and the materials used in the computer-based tests and the aforementioned paper-based tests were identical. The only difference between the two tests was the medium of delivery. As mentioned previously, the three participants had sufficient computer literacy to carry out the computer-based tests. Each of them was fully briefed about the procedures and settings of the tests. They were also allowed to try out the interface before carrying out the tests. The reading materials were displayed on the left side of the computer screen. The right side of the screen were divided into two sections. The upper part was a blank word document created for students to take notes when listening to the recordings. The lower part was another word document, the main writing section where the participants typed their essays. The computer-based tests were piloted with two first-year students in a university in Hong Kong to ensure that the interface is user-friendly and to try out the test design. Feedback from the two participants was collected to inform the main study. The eye tracker used in the study was Tobii TX300, and the eye tracking study consisted of the following four steps:

Step 1 Each participant went through a calibration procedure for eye fixations to ensure that the eye tracker can accurately keep track of participants' eye-movements during the test.

Step 2 Participants had about three minutes to skim the six reading passages before the listening part. Students could highlight the words and sentences they found important by selecting the desired part and pressing 1 on the keyboard. By pressing 2, they could undo the highlights. Participants then listened to a recording for approximately 12 min in each test. As with the paper-based tests, students could only listen to the recording once, whereas the reading passages were available to them throughout the tests. A word document displayed in the upper right corner of the screen was provided for participants to type notes from the recording.

Step 3 Participants wrote an essay in response to a given writing prompt in 1 h. The listening notes together with the six reading passages were available while they were writing their essays. In order to help the participants to keep track of time during the process, they were informed of the use of a timer staying minimized in the taskbar. They could check the time whenever needed.

Step 4 Participants took part in a stimulated-recall interview immediately after completion of an integrated writing test. Each participant was interviewed twice (one for the CIW and the other for the EIW). The interviews were conducted mainly in Cantonese, and each of them lasted approximately 20 min. The recorded eye movement videos were played as stimuli during the interviews. The procedure for the video-stimulated recall interview is summarized below.

During the tasks A researcher observed a participant doing a test and took notes on critical episodes observed on a secondary monitor of the participant's computer throughout the test process. 'Critical episodes' referred to any events that can signal the participant's use of a cognitive strategy while reading or writing (e.g., searching for information, reviewing or editing a sentence or paragraph, etc.). The episodes were noted down chronologically by indicating the locations (e.g., the specific word, phrase, sentence, or paragraph) and the actions performed by the participant (e.g., deletion, addition, revision, or pause).

During the replay of the video Before replaying the video, the participant was asked to describe the task s/he had just completed. The question was an icebreaker to help prepare the participant for the interview. Then the participant was invited to watch the task video with the researcher. The critical episodes noted down earlier were used to facilitate the interview.

The eye tracking data supplemented with the stimulated-recall interviews provided triangulating evidence that enabled us to interpret students' integrated writing performance from both the products (written production) and the processes (eye-movements). It was anticipated that successful writers would focus more or lingered longer on the relevant information, and that differences in integrated writing performance could be explained partly by the eye movement data and stimulated recall interviews.

Data analysis

The data collected from the paper-based integrated writing tests were sorted and entered into SPSS 24.0 for statistical analysis. Descriptive statistics (i.e., mean, standard deviation, skewness, and kurtosis) were calculated to examine the central tendencies, variation, and distributional properties of the data. Pearson product-moment correlation analysis was carried out to check the correlations between students' scores on the three discourse synthesis subscales (i.e., quotation, summarization, and connection), and the correlations between the three sub-scores and the holistic ratings for the CIW and the EIW respectively. We then performed a hierarchical regression analysis to determine the respective effect of the three discourse synthesis skills on students' performance on the CIW and the EIW. Students' scores on the discourse synthesis subscales formed the three independent variables, and the holistic rating was the dependent variable. In order to delve into the predictive contribution of the three discourse synthesis skills, we entered the data in different orders, resulting in six regression models for each hierarchical regression. Prerequisite assumptions including linear relationship, multivariate normality and multicollinearity were checked before regressions. The results showed that the values of variance inflation factor (VIF) for six independent variables used in regression models (i.e., three discourse synthesis variables in Chinese and three in English) ranged from 1.02 to 2.22. All values were substantially below the cutoff threshold for multicollinearity (i.e., $VIF = 10$), suggesting that the relationships among the variables did not impose severe impact on the regression analysis (Hair, Black, Babin, & Anderson, 2010).

Results

Paper-based integrated writing tests

Descriptive statistics

The descriptive statistics are shown in Table 1. In the CIW, the scores on summarization (C2) and connection (C3) were higher relative to quotation (C1). Whereas in the EIW, the scores on the three skills were generally low, with quotation (E1) being the highest, followed by summarization (E2), and connection (E3). The means of the holistic rating in the two tests differed, with 169.60 in the CIW and 111.96 in the EIW. In addition, all kurtosis and skewedness values were below 10, suggesting that the distributions of the variables were reasonably normal (Kline, 2016).

Correlational analysis

Pearson product-moment correlation analysis was performed to examine the relationships between variables. Table 2 shows that all the three skills had a significant positive correlation with students' overall integrated writing performance on both tests. In the CIW, summarization and- synthesis were both highly correlated with

Table 1 Descriptive statistics for discourse synthesis variables and the overall CIW and EIW performance (N = 145)

	Min	Max	<i>M</i>	<i>SD</i>	Skewness	Kurtosis
Quotation (C1)	0.00	8.50	3.05	2.62	0.16	− 1.38
Summarization (C2)	0.00	7.50	4.91	1.46	− 1.20	2.30
Connection (C3)	0.00	7.00	4.81	1.46	− 1.58	3.28
Subtotal (C1 + C2 + C3)	0.00	21.50	12.78	4.44	− 0.52	0.46
Quotation (E1)	0.00	6.50	2.82	2.27	− 0.23	− 1.63
Summarization (E2)	0.00	8.00	2.00	2.40	0.73	− 1.02
Connection (E3)	0.00	7.00	0.61	1.74	2.69	5.57
Subtotal (E1 + E2 + E3)	0.00	18.50	5.43	4.26	0.70	0.12
CIW (holistic)	33.00	258.50	169.61	42.66	− 0.99	1.00
EIW (holistic)	0.00	257.50	111.97	70.39	− 0.22	− 1.00

the holistic rating ($r = .72$), and quotation was moderately correlated ($r = .44$). In the EIW, summarization had the highest correlation with the holistic rating ($r = .57$), followed by quotation ($r = .40$) and synthesis ($r = .35$). Additionally, students' overall performance on the CIW was significantly correlated with their EIW performance ($r = .28$). We examined further how scores on the three discourse synthesis subscales in one test correlated with their counterparts in the other test. In the CIW, the scores on quotation (C1) and summarization (C2) both had a significant positive correlation with their counterparts in the EIW (E1 and E2), with a correlation coefficient $r = .17$ in quotation and $r = .23$ in summarization. In addition, quotation (C1) and summarization (C2) in the CIW were both positively correlated with the overall EIW scores. However, no significant correlation existed between the connection variables (C3) and (E3), and between the connection (C3) and the overall EIW scores.

Hierarchical regression analysis

The hierarchical regression analysis indicated that 63.6% of the variance in students' overall CIW performance was accounted for by the three discourse synthesis variables. As shown in Table 3, all the three predictors made significant contribution to the CIW scores, regardless of the order in which they were entered in the regression models. In step 1, connection and summarization made almost equal contribution by significantly accounting for 53% of the variance of the overall integrated writing performance, followed by quotation explaining 19.4% of the variance with $p < .001$. Table 4 provides the results of hierarchical regression analysis of the EIW. The three variables significantly accounted for 47.9% of the variance in the holistic ratings. Each discourse synthesis skill made significant contribution to the overall EIW performance, even though we controlled the effects of the other two skills. In step 1, summarization was the greatest predictor by explaining 32.9% of the score variance, followed by quotation (16.4%) and connection (12.9%), all of which were at $p < .001$ level. We delved further into the predictive relationship between the discourse synthesis skills in the CIW and students' overall EIW performance. Table 5

Table 2 bivariate correlations of discourse synthesis variables and scores on CIW and EIW

	Quotation (C1)	Summarization (C2)	Connection (C3)	Quotation (E1)	Summarization (E2)	Connection (E3)	Subtotal (C1 + C2 + C3)	Subtotal (E1 + E2 + E3)	CIW holistic	EIW holistic
Quotation (C1)	1									
Summarization (C2)	.36**	1								
Connection (C3)	.35**	.73**	1							
Quotation (E1)	.18*	.08	.01	1						
Summarization (E2)	.09	.24**	.15	.11	1					
Connection (E3)	.04	.12	.06	-.03	.40**	1				
Subtotal (C1 + C2 + C3)	.83**	.78**	.78**	.13	.18*	.08	1			
Subtotal (E1 + E2 + E3)	.16	.22**	.11	.58**	.78**	.61**	.21*	1		
CIW holistic	.44**	.73**	.73**	.11	.17*	.01	.74**	.16	1	
EIW holistic	.30**	.25**	.15	.41**	.57**	.36**	.31**	.69**	.29**	1

* $p < .05$; ** $p < .01$

Table 3 Hierarchical regression analysis predicting chinese integrated writing scores with discourse synthesis skills

	Step 1	Step 2	Step 3
<i>Model 1</i>			
β Quotation(C1)	7.168***	3.305***	2.676***
β Summarization (C2)		19.150***	11.147***
β Connection (C3)			11.478***
R ²	.194***	.566***	.636***
ΔR ²		.372***	.070***
<i>Model 2</i>			
β Quotation	7.168***	3.419***	2.676***
β Connection		19.184***	11.478***
β Summarization			11.147***
R ²	.194***	.571***	.636***
ΔR ²		.376***	.065***
<i>Model 3</i>			
β Summarization	21.308***	19.150***	11.147***
β Quotation		3.305***	2.676**
β Connection			11.478***
R ²	.530***	.566***	.636***
ΔR ²		.036***	.070***
<i>Model 4</i>			
β Summarization	21.308***	12.241***	11.147***
β Connection		12.370***	11.478***
β Quotation			2.676***
R ²	.530***	.613***	.636***
ΔR ²		.083***	.023***
<i>Model 5</i>			
β Connection	21.348***	19.539***	11.640***
β Quotation		3.087***	2.415***
β Summarization			11.306***
R ²	.532***	.566***	.636***
ΔR ²		.034***	.067***
<i>Model 6</i>			
β Connection	21.348***	12.370***	11.478***
β Summarization		12.241***	11.147***
β Quotation			2.676***
R ²	.532***	.613***	.636***
ΔR ²		.081***	.023***

** $p < .01$; *** $p < .001$

shows the cross-linguistic effects of Chinese discourse synthesis skills on English integrated writing. The regression analysis indicated that the scores on the three discourse synthesis skills in the CIW significantly accounted for 12.3% of the variance of the overall scores on the EIW. With respect to individual skills, summarization

Table 4 Hierarchical regression analysis predicting english integrated writing scores with discourse synthesis skills

	Step 1	Step 2	Step 3
<i>Model 1</i>			
β Quotation	12.578***	10.832***	11.282***
β Summarization		15.723***	13.480***
β Connection			7.684**
R ²	.164***	.449***	.479***
ΔR ²		.285***	.030**
<i>Model 2</i>			
β Quotation	12.578***	12.973***	11.282***
β Connection		15.139***	7.684***
β Summarization			13.480***
R ²	.164***	.304***	.479***
ΔR ²		.139***	.176***
<i>Model 3</i>			
β Summarization	16.794***	15.723***	13.480***
β Quotation		10.832***	11.282***
β Connection			7.684***
R ²	.329***	.449***	.479***
ΔR ²		.121***	.030***
<i>Model 4</i>			
β Summarization	16.794***	14.973***	13.480***
β Connection		6.365*	7.684***
β Quotation			11.282***
R ²	.329***	.350***	.479***
ΔR ²		.021*	.130***
<i>Model 5</i>			
β Connection	14.562***	15.139***	7.684***
β Quotation		12.973***	11.282***
β Summarization			13.480***
R ²	.129***	.304***	.479***
ΔR ²		.175***	.176***
<i>Model 6</i>			
β Connection	14.562***	6.365*	7.684***
β Summarization		14.973***	13.480***
β Quotation			11.282***
R ²	.129***	.350***	.479***
ΔR ²		.220***	.130***

* $p < .05$; ** $p < .01$; *** $p < .001$

and quotation both made positive contribution to the overall EIW scores at the 0.05 level, whereas synthesis did not ($p = .26$). In summary, the three discourse synthesis skills (i.e., quotation, summarization, and connection) were significant predictors of students' overall integrated writing performance in both L1 (Chinese) and

Table 5 Regression analysis predicting english integrated writing performance with scores on chinese discourse synthesis skills

	Unstandardized Coefficients		Standardized Coefficients	<i>t</i>	Sig.
	B	Standard error			
Quotation (C1)	6.889	2.292	.257	3.006	.003
Summarization (C2)	12.363	5.677	.256	2.178	.031
Connection (C3)	− 6.395	5.652	− .132	− 1.131	.260
The model	R ² = .123; F = 6.598***				

****p* < .001

L2 (English) contexts. In the CIW, summarization and connection were two largest contributors to the overall integrated writing performance. The predictive value of the three skills decreased in the EIW, with summarization being the largest contributor to the variance of overall writing quality.

Eye movement data

Based on the results of the computer-based tests, the three students were divided into three proficiency groups: (1) high performance in both tests (Student A); (2) high performance in the Chinese test, but low performance in the English test (Student C); (3) low performance in both tests (Student B). We defined areas of interest (AOI) in each reading passage based on its relevance to the writing topic. A reading passage may contain relevant AOIs and/or irrelevant AOIs. In the CIW, there were 10 relevant and three irrelevant AOIs. In the EIW, six relevant and three irrelevant AIOs were defined. The eye tracking data consisted of full recordings of the three participants’ eye movements while they were doing the tests. The data were analyzed in terms of Fixation duration, Fixation count, Visit duration, and Visit count, within individual AOIs and AOI groups. The results of data analysis were compared among the three students and reported below.

As shown in Tables 6 and 7, an obvious contrast between the high-performing and low-performing students was observed in the CIW. Student A and Student C, the two higher-achievers, spent over 83% and 95% of the time reading the texts relevant to the task, and 16% and 4% of the time on the irrelevant texts respectively. In contrast, Student B, the low-performer, spent 60% of the time on the irrelevant source texts, and only 39% on the relevant texts. Similar patterns recurred across the cases of Total visit duration, Fixation count, and Visit count. The two distinct and recurrent patterns confirmed the findings of the paper-based study that students’ approaches to source selection were associated with their integrated writing performance. The patterns also corroborated the assumption that high-performing writers are more capable of differentiating the relevant source materials from the irrelevant ones and will demonstrate a higher level of engagement with key information than their low-performing peers.

Table 6 Total fixation duration and total visit duration (in seconds)^a: CIW

AOI	Total fixation duration			Total visit duration		
	Student A	Student B	Student C	Student A	Student B	Student C
Rectangle 1	13.65	20.61	0.83	15.78	26.6	0.95
Rectangle 2 ^b	45.96	172.69	1.33	51.51	217.66	1.45
Rectangle 3	36.05	13.86	0.87	38.63	18.28	0.87
Rectangle 4	42.58	14	3.17	46.86	20.63	4.12
Rectangle 5	60.04	20.63	2.18	67.89	29.3	2.82
Rectangle 6	15.48	9.98	0.95	16.93	14.55	1.48
Rectangle 7	45.14	24.3	3.13	50.18	32.95	3.75
Rectangle 8	89.87	5.58	16.75	105.5	9.53	20.25
Rectangle 9	36.6	1.22	5.97	40.63	3.28	7.47
Rectangle 10 ^b	14.25	0.48	0	15.86	0.77	0
Rectangle 11 ^b	13.68	0.55	0.13	14.2	0.55	0.15
Rectangle 12	37.08	0.38	0.13	40.35	0.43	0.13
Rectangle 13	11.03	0.87	0	11.88	1.43	0
Sum	461.41	285.15	35.44	516.2	375.96	43.44
Relevant (%)	387.52 (83.99%)	111.43 (39.08%)	33.98 (95.88%)	434.63 (84.20%)	156.98 (41.75%)	41.84 (96.32%)
Irrelevant: 2, 10, 11 (%)	73.89 (16.01%)	173.72 (60.92%)	1.46 (4.12%)	81.57 (15.80%)	218.98 (58.25%)	1.6 (3.68%)

^aTotal fixation duration refers to the total amount of time an individual participant spent on an AOI. Total visit duration refers to the amount of time aggregated from individual visits to an AOI

^bIrrelevant AOI

Table 7 Fixation count and visit count (count)^a: CIW

AOI	Fixation count			Visit count		
	Student A	Student B	Student C	Student A	Student B	Student C
Rectangle 1	81	105	6	27	53	3
Rectangle 2 ^b	244	740	7	77	163	4
Rectangle 3	140	75	6	60	37	6
Rectangle 4	224	90	24	106	44	10
Rectangle 5	321	111	12	123	50	9
Rectangle 6	81	61	8	42	30	5
Rectangle 7	239	146	18	101	57	7
Rectangle 8	482	42	90	136	21	19
Rectangle 9	198	8	33	88	5	8
Rectangle 10 ^b	71	5	0	33	2	0
Rectangle 11 ^b	55	4	2	33	4	1
Rectangle 12	170	3	1	67	2	1
Rectangle 13	49	7	0	22	4	0
Sum	2355	1397	207	915	472	73
Relevant (%)	1985 (84.29%)	648 (46.39%)	198 (95.65%)	772 (84.37%)	303 (64.19%)	68 (93.15%)
Irrelevant: 2, 10, 11 (%)	370 (15.71%)	749 (53.61%)	9 (4.35%)	143 (15.63%)	169 (35.81%)	5 (6.85%)

^aFixation count represents the frequency a participant fixates on an AOI. Visit count represents the frequency a participant visits an AOI

^bIrrelevant AOI

The subsequent stimulated-recall interview indicated that Student B was hesitant if he should use Source 1, represented by Rectangle 1 to 3, or Source 3, represented by Rectangle 4 to 6, to start the first paragraph of his essay. At first, he was planning to relate Source 3, a bar chart illustrating an aging population in Hong Kong, to yielding seats on public transportation. He then changed his mind and decided to describe all the school events listed in Source 1 consecutively. Student B explained that he did so because these events were all important. They could “inform classmates about what activities are helpful in building understanding of yielding seats...things related to priority seats”, and “because usually activities can tell people what priority seats are about”. Clearly, Student B’s interpretation of the writing task deviated from the task instruction which asked writers to synthesize contrasting perspectives on priority seats, rather than conceptualizing priority seats. The two high performers, Student A and Student C, on the other hand, showed a greater awareness of task requirements. When asked to explain the thinking behind their selection of source materials, both of them referred to the writing prompt. Scrolling up and down of the reading materials was observed repeatedly in the case of Student A in different phases of the

test. In the interview, he explained that he was “looking for information appropriate to the task” and that he did so to make sure of his essay responding to the task. Student C took notes in the first three minutes of the test while reading the six passages. He explained that “In fact, my understanding (of the topic) was formed after reading through the texts, listening to the recording. I synthesized the information, put together the materials and used my words to write them up”. The interview data revealed that the two high performers had a clear goal in mind in the pre-writing stage and throughout the writing process. A clearer understanding of task requirements contributed to strategic synthesis of source materials and in turn better integrated writing performance.

In the English task, a similar pattern was observed across the three participants. As shown in Tables 8 and 9, regardless of achievement level, all of them spent considerably more time on the relevant source texts, varying from 85% to 92% of the Total fixation duration, relative to the time spent on the irrelevant texts, which was in the range of 7% to 14%. The consistent pattern also existed in Total visit duration, Fixation count, and Visit count. Unlike what we observed in the Chinese task, there seemed to be no obvious difference among the three participants in terms of their approaches to source selection, even though their test performance differed.

The interview data helped us to probe into this phenomenon. Student A fixated considerably longer on Source 3 and 4 (i.e., Rectangle 4 and 5). In the interview, he explained that he was thinking how to start the paragraph regarding the negative and positive aspects of living organ donation and pondering “whether there was stuff I could put in the essay”. In fact, the student did not think reading six written texts was too much for him to handle; however, he admitted that “I am still puzzled by some sentences after reading through the passages”. Frequent long pauses were observed in the case of Students C who outperformed his peers in the CIW. In the EIW, on the other hand, Student C seemed to be constrained by his language proficiency. The student attributed the frequent interruptions observed to extra cognitive processes of translating vocabulary from Chinese to English and retrieving appropriate wording, as the quotation below indicated:

I usually read Chinese books like novels. I have a bunch of Chinese vocabulary in my mind, but I can hardly put the words in English. I don't know how to translate them. If I am writing in Chinese, I can write anything that crossed my mind freely. When writing in English, if something occurred to me, and I don't know how to phrase it, I will halt there and ponder how to write.

The higher level of cognitive demand in L2 integrated writing was also manifested in verbatim source use as exhibited by Student A who directly copied some phrases and sentences from the source texts in both tests, but for very different reasons. In the CIW, Student A quoted a concluding sentence from Source 2 with minor modifications made himself. The quotation below illustrated why Student A chose to use the source text nearly verbatim rather than writing in his own words.

Q: We can tell from the color that how many words you copied and pasted from Source 2. Could you explain why you did so?

A: Because the sentence from Source 2 helped me conclude the content.

Table 8 Total fixation duration and total visit duration (in seconds): EIW

AOI	Total Fixation Duration			Total visit duration		
	Student A	Student B	Student C	Student A	Student B	Student C
Rectangle 1	88.61	10.12	17.03	95.35	11.72	19.35
Rectangle 2	93.64	60.06	50.03	109.87	79.77	61.81
Rectangle 3 ^a	40.45	8.7	7.68	43.18	10.32	9.17
Rectangle 4	517.74	41.46	75.07	579.88	53.21	91.12
Rectangle 5	342.51	45.76	91.02	387.5	67.04	124.73
Rectangle 6	50.83	14.9	2.45	54.03	17.21	2.83
Rectangle 7	53.21	2.03	0.87	57.73	2.25	1.2
Rectangle 8 ^a	26.35	0	0	28.23	0	0
Rectangle 9 ^a	29.3	10.32	33.55	32.2	14.75	41.98
Sum	1242.64	193.35	277.7	1387.97	256.27	352.19
Relevant (%)	1146.54 (92.27%)	174.33 (90.16%)	236.47 (85.15%)	1284.36 (92.54%)	231.2 (90.22%)	301.04 (85.48%)
Irrelevant: 3, 8, 9 (%)	96.1 (7.73%)	19.02 (9.84%)	41.23 (14.85%)	103.61 (7.46%)	25.07 (9.78%)	51.15 (14.52%)

^aIrrelevant AOI

Table 9 Fixation count and visit count (count): EIW

AOI	Fixation count			Visit count		
	Student A	Student B	Student C	Student A	Student B	Student C
Rectangle 1	363	45	81	126	23	46
Rectangle 2	499	313	238	164	98	71
Rectangle 3 ^a	188	47	41	115	27	21
Rectangle 4	2255	207	301	608	78	57
Rectangle 5	1513	267	473	436	103	109
Rectangle 6	222	57	11	108	26	5
Rectangle 7	236	12	9	86	8	6
Rectangle 8 ^a	102	0	0	38	0	0
Rectangle 9 ^a	111	68	159	22	23	13
Sum	5489	1016	1313	1703	386	328
Relevant (%)	5088 (92.69%)	901 (88.68%)	1113 (84.77%)	1528 (89.72%)	336 (87.05%)	294 (89.63%)
Irrelevant: 3, 8, 9 (%)	401 (7.31%)	115 (11.32%)	200 (15.23%)	175 (10.28%)	50 (12.95%)	34 (10.37%)

^aIrrelevant AOI

Q: Why did you choose to copy and paste from Source 2?

A: To save time.

Q: Would it be possible for you to write a better sentence than the one you copied if you write in your own words?

A: Yes. I think I can make it.

It is clearly that Student A put efficiency above originality when copying the sentence from Source 2; nevertheless, he was quite confident that he could have done better if he had written in his own words. In the EIW, on the contrary, almost half of the length of the third paragraph came exactly from Source 3. In the subsequent interview, instead of referring to that as a time-saving strategy, Student A indicated that the sentences he copied were needed for explaining why people were hesitant about living organ donation.

Q: Why did you copy the whole sentence from Source 3?

A: Hmm, because this sentence mentioned the impact?

Q: Do you mean the impact on living donors?

A: People don't know what the impact is. This also explained why there are so few people here.

Q: So this explained further the negative impact?

A: No.

Q: No? What is it?

A: It just talked about why there are so few living donors.

The above quotations revealed two contrasting attitudes toward verbatim source use. When writing in L1, Student A used the source materials strategically,

considering verbatim source use as a time-saving strategy, whereas in L2 writing, the student turned to the source text for language support to make his point. Limited language resources resulted in over-reliance on source texts, thereby a substantial proportion of inappropriate verbatim source use.

Discussion

The present study examined the relationships between discourse synthesis skills (i.e., quotation, summarization, and connection) and the overall integrated writing performance in both Chinese, students' L1, and English, their L2. Results of the integrated writing tests together with the eye-movement data and stimulated-interviews unveiled the differing effects of discourse synthesis skills on L1 and L2 integrated writing and the cross-linguistic effect of these skills. These findings will be discussed in detail below.

Discourse synthesis skills and cross-linguistic effect

Hierarchical regression analysis was performed to answer the first research question: what are the effects of discourse synthesis skills on students' integrated writing performance? Our results concurred with previous studies that discourse synthesis skills are significant indicators of integrated writing performance (Keck, 2014; Plakans & Gebriel, 2013, 2017), explaining up to 63.6% of the variance in overall writing performance in L1 and 47.9% in L2. In the CIW, summarization and connection made nearly equal contribution and altogether accounted for 61.3% of the variance of the overall writing quality. In the EIW, even though summarization and connection remained significant predictors, the joint predictive power reduced to 35%. Summarization in concert with quotation had greater contribution, explaining 44.9% of the variance of the overall English writing quality. The consistent predictions substantiate the validity of the two writing tasks that elicited underlying discourse synthesis skills in integrated writing.

The present study extended previous knowledge by investigating the cross-linguistic effect of the discourse synthesis skills from L1 to L2 in order to answer the second research question: what is the relationship between students' L1 and L2 discourse synthesis performance, and is there a cross-linguistic effect of L1 discourse synthesis skills on L2 writing? The regression analysis indicated that 12.3% of the overall EIW performance was contributed by Chinese discourse synthesis skills. Scores on quotation and summarization in the CIW were both significant predictors of students' overall EIW performance, even though connection score was not. The finding suggests that development of the abilities to select and summarize information from sources in L1 may facilitate the overall L2 integrated writing performance (Savage et al., 2017; Schoonen et al., 2003; Sparks, 2012; Sparks et al., 2009, 2012). Nevertheless, a decrease in predictive strength of discourse synthesis was observed (63.6% in L1 and 47.9% in L2). It is particularly noteworthy in the between-language case (12.3% in L1 to L2). The striking reduction might be associated with the

smaller contribution made by the Chinese connection variable to the overall EIW performance. Although the Chinese discourse synthesis scores altogether contributed to 12.3% of the overall English integrated writing performance, the connection variable (C3) alone was not significantly correlated with the overall EIW quality. A possible explanation to the weaker predictive strength of the connection variable in the L2 assessment and the small cross-linguistic effect of L1 on L2 is that students' connection skill in English had not yet developed to the extent as it was in Chinese. The connection skill requires complex cognitive operations to link information across different modalities (i.e., reading and listening) and across multiple texts. Under timed conditions, students could select and summarize relevant information from individual source materials, whereas they might still find it difficult to draw together the key information inter-textually with their second language.

Task representation and language proficiency in integrated writing

Despite the cross-linguistic effects of discourse synthesis skills from L1 to L2, findings of the computer-based tests and stimulated recall interviews revealed that students' approaches to discourse synthesis differed in L1 and L2 assessments. In the CIW, contrasting patterns were observed between the high-performing and the low-performing group, suggesting that eye movements are reflective of performance on integrated writing assessments (Bax & Chan, 2016; Yu et al., 2017). The stimulated recall interviews revealed that students' engagement with source texts was closely associated with their representations of the writing task in the pre-writing stage where metacognitive control over the planning, evaluation, and revision processes has a significant role to play (Whalen & Ménard, 1995; Yang & Plakans, 2012). Misinterpretation of task requirements caused inappropriate selection of source texts, thereby inclusion of irrelevant information as in the case of Student B who fixated considerably longer on the irrelevant text than his high-performing counterparts.

On the contrary, having a clear and precise goal in mind throughout the writing process contributed to strategic selection of source materials, enhanced awareness of discourse synthesis, and in turn better integrated writing performance. The results suggest that presented with the same L1 integrated writing task, students' interpretation of task requirements might vary considerably, thereby affecting the quality of their written products. The significance of task representation in integrated writing assessment has also been highlighted in previous studies. The formation of task representation involves an interpretive process that would evolve throughout the writing process; writing from multiple sources makes the interpretive process more challenging (Plakans, 2010). Researchers have also pointed out that task representation affects not only the product but also the process of integrated writing (Ruiz-Funes, 2001; Wolfersberger, 2013) and therefore, more attention should be paid to metacognitive control to guide students to check, evaluate, and regulate their thinking in the writing process (Ruiz-Funes, 2001).

Notwithstanding the differences in the overall EIW scores, the eye-movement data collected from the three students did not indicate noticeable variations in their

abilities to differentiate relevant sources from irrelevant ones, as observed in the CIW. The inconsistent findings emerging from the CIW and the EIW contribute to Keck's (2014) study by indicating that similar practices of text selection and summarization may not lead to equivalent writing quality. The stimulated recall interviews further unveiled that students had extra hurdles to clear when writing in L2 (Schoonen et al., 2003). As discussed previously, the employment of discourse synthesis skills in L1 writing is influenced by the writing goals students set in the pre-writing stage. When writing in their second language, the students seemed to be less articulate about their writing goals and tended to treat the source materials as unconnected. Rarely did they show an awareness of connecting ideas from different sources as observed in the CIW. This may explain why the predictive value of the discourse synthesis skills in the EIW was smaller than that in the CIW, particularly in the case of connection (Asención Delaney, 2008; Whalen & Ménard, 1995).

Linguistic constraints were also evident in the improper verbatim use of source texts as exhibited by Student A. Writing from sources in a second language involves not only integration of source materials but synergy of writer's first and second language (Plakans & Gebril, 2012). Writers who effectively selected, organized, and connected information from multiple texts tended to avoid patchwriting and direct copy and pasting from source texts (Cumming et al., 2005). Low performers who are lacking language resources may turn to source texts for language support, particularly when being pressured by time constraints in a test setting. Students might focus mainly on how to complete the writing tasks and over-rely on source materials to compensate for limited language resources, thus overlooking the risk of plagiarism inherent in verbatim use of sources. Plakans and Gebril (2012, 2013) have indicated that higher tendency for verbatim use of source texts is associated with weaker discourse synthesis performance. This may explain why discourse synthesis skills were less predictive of L2 integrated writing quality. The differences existed between L1 and L2 integrated writing also suggest that long fixations observed in L2 writing processes should be treated with caution. They did not necessarily represent engagement with source texts as observed in the case of L1 writing. Instead, long fixations might be reflective of students' struggle to comprehend source materials or every effort they made to get useful information from a source text. It is anticipated that development in L2 proficiency will gradually diminish the influence of linguistic constraints, thereby increasing the predictive strength of discourse synthesis on L2 integrated writing.

Conclusions

The present study confirmed the decisive role of discourse synthesis skills in L1 and L2 integrated writing assessments. Several implications can be drawn from this study. The results of hierarchical regression analysis provide evidence for construct validity, indicating that the three discourse synthesis skills (i.e., quotation, summarization, and connection) are significant constructs elicited by both L1 and L2 integrated writing assessments. Assessment designers might incorporate the three skills into assessment tasks and make them explicit in assessment requirements. The

use of fine-grained criteria and level descriptors with regard to discourse synthesis skills could help inform test-takers of their strengths and weaknesses and provide specific feedback for improvement in integrated writing. The findings also offer some insights into integrated writing instruction. Language teachers might adopt skill-based approaches to enhance students' ability to select, summarize, and connect information from multiple sources that present different or sometimes even conflicting perspectives. The decreasing effect of connection on L2 integrated writing assessment suggest that cross-text synthesis might be the most challenging part when students write from sources in their second language. Considering the transferability of discourse synthesis skills from L1 to L2, schools may regard discourse synthesis as cross-curricular skills and encourage Chinese and English language teachers to collaborate in designing integrated writing curriculum, with particular attention being paid to the ability to connect ideas from multiple source materials. Given that integrated writing is an integral part of the secondary education curriculum in Hong Kong, teachers may create authentic language environments by engaging students with language tasks that involves the employment of different language skills in everyday classrooms. In light of the overarching role of metacognition in discourse synthesis, incorporation of self-regulation strategies, such as goal setting, monitoring and evaluating writing plans and processes would be helpful in focusing students on information important to their written production and in facilitating strategic selection of source materials. It is also necessary to raise students' awareness of the boundary between appropriate verbatim source use and plagiarism to help them engage better in academic discourse.

Several limitations must be acknowledged. First, the results of the study support the significance of discourse synthesis skills in L1 and L2 integrated writing. However, we cannot infer that the practice of source integration is the same in Chinese and English. There might be rhetorical differences resulting from language conventions or writing instruction. A further exploration is needed to verify this assumption. Text analysis of written products would provide illuminating insights into the differences and/or similarities in students' integration practice in the two languages. Moreover, the present study focused specifically on students' ability to select, summarize, and synthesize source materials. Even though the importance of the three skills and the impact of language proficiency on discourse synthesis performance have been substantiated in the present study and the literature, it is still possible that other factors might influence integrated writing performance. Future studies may explore whether differences in the number of years studying English and exposure to English outside of schooling would account for differences in L2 discourse synthesis skills and the overall L2 integrated writing performance. A further exploration of L2 learning experience and its relationships to discourse synthesis could advance our understanding of integrated writing skills and shed light on writing instruction. In addition, the computer-based tests were conducted with a very small sample given the time-consuming and intensive nature of the process of data collection in the tests. The generalizability of the results should be treated with caution. The patterns emerged from the eye-movement data and the stimulated-recall interviews indeed lent support to the results of the paper-based assessments. Collecting data from a larger sample will

allow in-depth statistical analysis that helps make eye movement a more reliable source of data to explain differences in integrated writing performance.

Funding The study reported in this article was funded by the Research Grants Council of Hong Kong (Reference No. 15640416).

Appendix: Analytic integrated writing rubrics

(The Chinese integrated writing marking rubrics are consistent with the English rubrics.)

1. Quotation

Marks	Level descriptor
9–10	Important information is precisely quoted from the readings and/or the recording (i.e., no similar information can be summarized)
7–8	Important information is precisely quoted from the readings and/or the recording, but a small part of the information quoted needs to be summarized
5–6	Important information is quoted from the readings and/or the recording, but some of the information quoted is inappropriate (i.e., it is not important or needs to be summarized)
3–4	Information is quoted from the readings and/or the recording, but at least half of the information quoted is inappropriate (i.e., it is not important or needs to be summarized)
1–2	Most of the content is directly copied from the readings and/or from the recording. The writer made no attempt to distinguish key information from the source materials
0	No attempt was made to quote the source materials

2. Summarization (single source text)

Marks	Level descriptor
9–10	Important information presented in the readings and/or from the recording is comprehensively and concisely summarized
7–8	Important information presented in the readings and/or from the recording is purposefully summarized, but the expression is not concise
5–6	Some of the important information presented in the readings and/or from the recording is summarized while some key points are missing
3–4	Information presented in the readings and/or from the recording is summarized, but the selection of information is unbalanced with many key points missing
1–2	The writer rarely summarized the information. Most of the content is merely a restatement of the information presented in the readings or from the recording
0	No attempt was made to summarize the source materials

3. Connection (multiple source texts)

Marks Level descriptor

9–10	Differing perspectives (from the readings and the recording) are comprehensively and concisely synthesized
7–8	Differing perspectives (from the readings or the recording) are clearly synthesized
5–6	Differing perspectives (from the readings or the recording) are synthesized, but the content is not sufficient
3–4	An attempt to synthesize differing perspectives (from the readings or the recording) is made, but most of the content is not reasonable
1–2	An attempt to synthesize differing perspectives (from the readings or the recording) is not evident. Only a single perspective (for/against) is present
0	No attempt was made to synthesize information from the source materials

References

- Asención Delaney, Y. (2008). Investigating the reading-to-write construct. *Journal of English for Academic Purposes*, 7(3), 140–150. <https://doi.org/10.1016/j.jeap.2008.04.001>.
- Bax, S., & Chan, S. H. C. (2016). *Researching the cognitive validity of GEPT high-intermediate and advanced reading: An eye tracking and stimulated recall study*. Retrieved from <https://www.lttc.ntu.edu.tw/lttc-gept-grants/RReport/RG07.pdf>.
- Cheong, C.M., Zhu, X., & Liao, X. (2018). Differences between the relationship of L1 learners' performance in integrated writing with both independent listening and independent reading cognitive skills. *Reading and Writing*, 31(4), 779–811.
- Cheong, C. M., Zhu, X., Li, G. Y., & Wen, H. (2019). Effects of intertextual processing on L2 integrated writing. *Journal of Second Language Writing*, 44, 63–75.
- Cumming, A., Kantor, R., Baba, K., Erdosy, U., Eouanzoui, K., & James, M. (2005). Differences in written discourse in independent and integrated prototype tasks for next generation TOEFL. *Assessing Writing*, 10(1), 5–43. <https://doi.org/10.1016/j.asw.2005.02.001>.
- Cumming, A., Lai, C., & Cho, H. (2016). Students' writing from sources for academic purposes: A synthesis of recent research. *Journal of English for Academic Purposes*, 23, 47–58. <https://doi.org/10.1016/j.jeap.2016.06.002>.
- Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research*, 49(2), 222–251. <https://doi.org/10.3102/00346543049002222>.
- Curriculum Development Council, & Hong Kong Examinations and Assessment Authority. (2007a). *English language curriculum and assessment guide (Secondary 4–6)*. Hong Kong.
- Curriculum Development Council, & Hong Kong Examinations and Assessment Authority. (2007b). 中國語文課程及評估指引: 中四至中六 (*Chinese language curriculum and assessment guide: Secondary 4–6*) Hong Kong.
- Deygers, B., Van den Branden, K., & Peters, E. (2017). Checking assumed proficiency: Comparing L1 and L2 performance on a university entrance test. *Assessing Writing*, 32, 43–56. <https://doi.org/10.1016/j.asw.2016.12.005>.
- Education Bureau. (2010). *Enriching our language environment, realising our vision: Fine-tuning of medium of instruction for secondary schools*. Hong Kong: The Government of the Hong Kong Special Administrative Region Retrieved from https://www.edb.gov.hk/attachment/en/edu-system/primary-secondary/applicable-to-secondary/moi/moi_booklet-eng-17apr2010.pdf.
- Grabe, W. (2001). Reading-writing relations: Theoretical perspectives and instructional practices. In D. Belcher & A. Hirvela (Eds.), *Linking literacies: Perspectives on L2 reading-writing connections* (pp. 15–47). Ann Arbor, MI: The University of Michigan Press.

- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Upper Saddle River, NJ: Pearson.
- Hidi, S., & Anderson, V. (1986). Producing written summaries: Task demands, cognitive operations, and implications for instruction. *Review of Educational Research*, 56(4), 473–493. <https://doi.org/10.3102/00346543056004473>.
- Hong Kong Examinations and Assessment Authority. (2005). 香港中文會考中國語文科水平參照等級描述及示例 (*Hong Kong Certificate of Education Examination Chinese language level descriptors and exemplars*). Hong Kong: Government Printer.
- Hong Kong Examinations and Assessment Authority. (2010). *HKCEE Chinese language examination report and question papers*. Hong Kong: Hong Kong Examinations and Assessment Authority.
- Hong Kong Examinations and Assessment Authority. (2012). *HKDSE Chinese language examination report and question papers*. Hong Kong: Hong Kong Examinations and Assessment Authority.
- Hong Kong Examinations and Assessment Authority. (2017). *2017 Hong Kong Diploma of Secondary Education Examination question papers (with marking schemes and candidates' performance)*. Hong Kong: Hong Kong Examinations and Assessment Authority.
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87(4), 329–354. <https://doi.org/10.1037/0033-295X.87.4.329>.
- Keck, C. (2014). Copying, paraphrasing, and academic writing development: A re-examination of L1 and L2 summarization practices. *Journal of Second Language Writing*, 25, 4–22. <https://doi.org/10.1016/j.jslw.2014.05.005>.
- Keung, Y.-C., & Ho, C. S.-H. (2009). Transfer of reading-related cognitive skills in learning to read Chinese (L1) and English (L2) among Chinese elementary school children. *Contemporary Educational Psychology*, 34(2), 103–112. <https://doi.org/10.1016/j.cedpsych.2008.11.001>.
- Kim, Y.-S. G., & Piper, B. (2019). Cross-language transfer of reading skills: An empirical investigation of bidirectionality and the influence of instructional environments. *Reading and Writing*, 32(4), 839–871. <https://doi.org/10.1007/s11145-018-9889-7>.
- Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). New York, NY: Guilford Press.
- Leki, I., Cumming, A., & Silva, T. (2008). *A synthesis of research on second language writing in English*. New York, NY: Routledge.
- Plakans, L. (2009). Discourse synthesis in integrated second language writing assessment. *Language Testing*, 26(4), 561–587. <https://doi.org/10.1177/0265532209340192>.
- Plakans, L. (2010). Independent vs. Integrated writing tasks: A comparison of task representation. *TESOL Quarterly*, 44(1), 185–194. <https://doi.org/10.5054/tq.2010.215251>.
- Plakans, L., & Gebril, A. (2012). A close investigation into source use in integrated second language writing tasks. *Assessing Writing*, 17(1), 18–34. <https://doi.org/10.1016/j.asw.2011.09.002>.
- Plakans, L., & Gebril, A. (2013). Using multiple texts in an integrated writing assessment: Source text use as a predictor of score. *Journal of Second Language Writing*, 22(3), 217–230. <https://doi.org/10.1016/j.jslw.2013.02.003>.
- Plakans, L., & Gebril, A. (2017). Exploring the relationship of organization and connection with scores in integrated writing assessment. *Assessing Writing*, 31, 98–112. <https://doi.org/10.1016/j.asw.2016.08.005>.
- Ruiz-Funes, M. (2001). Task representation in foreign language reading-to-write. *Foreign Language Annals*, 34(3), 226–234. <https://doi.org/10.1111/j.1944-9720.2001.tb02404.x>.
- Savage, R., Kozakewich, M., Genesee, F., Erdos, C., & Haigh, C. (2017). Predicting writing development in dual language instructional contexts: Exploring cross-linguistic relationships. *Developmental Science*. <https://doi.org/10.1111/desc.12406>.
- Sawaki, Y., Quinlan, T., & Lee, Y.-W. (2013). Understanding learner strengths and weaknesses: Assessing performance on an integrated writing task. *Language Assessment Quarterly*, 10(1), 73–95. <https://doi.org/10.1080/15434303.2011.633305>.
- Schoonen, R., van Gelderen, A., de Glopper, K., Hulstijn, J., Simis, A., Snellings, P., et al. (2003). First language and second language writing: The role of linguistic knowledge, speed of processing, and metacognitive knowledge. *Language Learning*, 53(1), 165–202. <https://doi.org/10.1111/1467-9922.00213>.
- Segev-Miller, R. (2007). Cognitive processes in discourse synthesis: The case of intertextual processing strategies. In M. Torrance, L. Van Waes, & D. Galbraith (Eds.), *Writing and cognition: Research and applications* (pp. 231–250). Amsterdam: Elsevier.

- Shum, K. K., Ho, C. S., Siegel, L. S., & Au, T. K. (2016). First-language longitudinal predictors of second-language literacy in young L2 learners. *Reading Research Quarterly*, 51(3), 323–344. <https://doi.org/10.1002/rrq.139>.
- Sparks, R. L. (2012). Individual differences in L2 learning and long-term L1–L2 relationships. *Language Learning*, 62(s2), 5–27. <https://doi.org/10.1111/j.1467-9922.2012.00704.x>.
- Sparks, R. L., Patton, J., Ganschow, L., & Humbach, N. (2009). Long-term crosslinguistic transfer of skills from L1 to L2. *Language Learning*, 59(1), 203–243. <https://doi.org/10.1111/j.1467-9922.2009.00504.x>.
- Sparks, R. L., Patton, J., Ganschow, L., & Humbach, N. (2012). Do L1 reading achievement and L1 print exposure contribute to the prediction of L2 proficiency? *Language Learning*, 62(2), 473–505. <https://doi.org/10.1111/j.1467-9922.2012.00694.x>.
- Sparks, R. L., Patton, J., & Luebbbers, J. (2019). Individual differences in L2 achievement mirror individual differences in L1 skills and L2 aptitude: Crosslinguistic transfer of L1 to L2 skills. *Foreign Language Annals*. <https://doi.org/10.1111/flan.12390>.
- Spivey, N. N., & King, J. R. (1989). Readers as writers composing from sources. *Reading Research Quarterly*, 24(1), 7–26.
- Wang, M., Perfetti, C. A., & Liu, Y. (2005). Chinese-English biliteracy acquisition: Cross-language and writing system transfer. *Cognition*, 97(1), 67–88. <https://doi.org/10.1016/j.cognition.2004.10.001>.
- Whalen, K., & Ménard, N. (1995). L1 and L2 writers' strategic and linguistic knowledge: A model of multiple-level discourse processing. *Language Learning*, 45(3), 381–418. <https://doi.org/10.1111/j.1467-1770.1995.tb00447.x>.
- Wolfersberger, M. (2013). Refining the construct of classroom-based writing-from-readings assessment: The role of task representation. *Language Assessment Quarterly*, 10(1), 49–72. <https://doi.org/10.1080/15434303.2012.750661>.
- Yang, H.-C., & Plakans, L. (2012). Second language writers' strategy use and performance on an integrated reading-listening-writing task. *TESOL Quarterly*, 46(1), 80–103. <https://doi.org/10.1002/tesq.6>.
- Yu, G. (2009). The shifting sands in the effects of source text summarizability on summary writing. *Assessing Writing*, 14(2), 116–137.
- Yu, G. (2013). The use of summarization tasks: Some lexical and conceptual analyses. *Language Assessment Quarterly*, 10(1), 96–109.
- Yu, G., He, L., & Isaacs, T. (2017). The cognitive processes of taking IELTS Academic Writing Task 1: An eye-tracking study. Retrieved from https://www.ielts.org/teaching-and-research/research-reports/ielts_online_rr_2017-2
- Zhu, X. (2005). *A study of setting standards of Chinese Language assessment in reading, writing and integrated skills (Unpublished technical report, project no: CD/C/13-0309)*. Singapore: National Institute of Education, Nanyang Technological University.
- Zhu, X., Li, X., Yu, G., Cheong, C. M., & Liao, X. (2016). Exploring the relationships between independent listening and listening-reading-writing tasks in Chinese language testing: Toward a better understanding of the construct underlying integrated writing tasks. *Language Assessment Quarterly*, 13(3), 167–185