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

Xinhua Zhu[a]

Xueyan Li[b]

Guoxing Yu[c]

Choo Mui Cheong[d]

Xian Liao[a]

[a]The Hong Kong Polytechnic University, Hong Kong SAR, People's Republic of China; [b]Central China Normal University, Wuhan, People's Republic of China; [c]University of Bristol, Bristol, United Kingdom; [d]The University of Hong Kong, Hong Kong SAR, People's Republic of China

Abstract

Integrated assessment tasks have been increasingly used in language tests, but the underlying constructs of integrated tasks remain elusive. This study aimed to improve understanding of the construct of integrated writing tasks in Chinese Language examinations in Hong Kong by looking at the language competences measured in the Listening-Reading-Writing Task and how they relate to the outcome of the Independent Listening Task. The performance of 226 native Chinese Secondary Five students on both tasks were subject to correlation analysis, joint factor analysis, and regression analysis. It was found that the students' performance in the Independent Listening Task and the Listening-Reading-Writing Task was statistically significantly correlated, but the two tasks did not seem to have common factors as shown in the joint factor analysis. The indicators of elaboration, evaluation, and creation in the Independent Listening Task were significantly correlated with multiple indicators in the Listening- Reading-Writing Task, and evaluation and creation together explained 8.9% of the variance in the total score of the Listening-Reading-Writing Task. The findings support the framework (i.e., the "four pillars" of integrated writing competence) applied in public examinations in Hong Kong. They also imply that the two types of writing tasks are complementary in the assessment of Chinese Language competence.

Introduction

The independent task in language tests is an entrenched approach used in large-scale standardized language testing for first and second languages. It requires test takers to produce language in either spoken or written form based on prior knowledge or experience (Barkaoui, Brooks, Swain, & Lapkin, 2013; Brown, Iwashita, & McNamara, 2005). Examining the validity of independent writing tasks, Cho (2003) commented that they may fail to adequately measure the writing construct because the writing produced lacks authenticity. Furthermore, these tasks do not provide input in any form and assume that test takers possess relevant background knowledge about the topic, thus contaminating the construct validity of writing tasks (Gebril, 2009; Gebril & Plakans, 2009; Weigle, 2004). Unlike independent tasks, which are decontextualized, integrated tasks require test takers to listen to and/or read sources and produce an appropriate oral or written output. Because of their theoretical validity, practical authenticity, and fairness in testing, integrated tasks are increasingly getting popular in international and regional language tests (Barkaoui et al., 2013; Gebril, 2009). Below we present a brief review of four key areas of research into integrated writing tasks: (a) the relationships between test-takers' performance on independent and integrated tasks; (b) the relationship between test-takers' language proficiency, especially their skills in comprehending source texts, and their performance in integrated tasks; (c) the discourse features of integrated writings; and (d) test-taking process (e.g., discourse synthesis).

Literature review

A number of studies have compared students' performance in independent and integrated tasks. Some suggested that independent and integrated writing tasks measure a similar (or the same) construct (e.g., Cumming et al., 2005; Gebril, 2006, 2009, 2010; Lee, 2006; Lewkowicz, 1994; Watanabe, 2001); others argued that the two tasks possess somewhat different constructs (e.g., Guo, Crossely, & McNamara, 2013). Looking at the product of integrated tasks, some studies have found more complicated functional and textual organization in integrated tasks than in independent tasks (Brown et al., 2005); others have revealed that test takers tended to include more ideas, which are often less developed, in their integrated writings than in independent writing responses (Lewkowicz, 1994). In a study of various independent and integrated TOEFL iBT tasks, Sawaki, Stricker, and Oranje (2009) showed that test-takers' performance in integrated speaking and writing tasks are highly associated with their independent speaking and writing skills, respectively. These studies seemed to suggest that while students' performance may be similar across independent and integrated tasks, there are differences in their underlying constructs. However, these findings do not illuminate the relationships between the underlying constructs of the independent and integrated tasks.

Integrated writing involves the use of sources that are not available in independent writing (Plakans & Gebril, 2013). A number of studies examined students' comprehension and use of such sources in their integrated writing. Students with different levels of language proficiency demonstrate varied use of source materials during integrated writing. The quantity of content from the source materials used in the final product of integrated tasks increases with test-takers' language proficiency (Brown et al., 2005), thus indicating that the comprehension and the interpretation of the source materials contribute to the construct of integrated writing assessment (Esmaeili, 2002; Gebril & Plakans, 2013; Plakans, 2009a; Sawaki, Quinlan, & Lee, 2013; Wolfersberger, 2013). When using the source materials, high-scoring test takers also exhibit higher-order thinking skills, while lower-scoring test takers depend heavily on copying words and phrases directly from source materials (Plakans & Gebril, 2013). In listening-writing or listening-speaking integrated tasks, the listening process is even more complex than the comprehension of written text in reading-writing integrated tasks. For instance, note taking in classroom context, an essential activity in listening-writing tasks, would involve a series of subprocesses, including selecting, constructing, and transforming the source (Peverly et al., 2007).

Furthermore, Cumming (2013) rightly pointed out that there might be a threshold for L2 learners to comprehend source materials to successfully complete integrated tasks. By and large, we observed the scarcity of research into the comprehension of listening materials in integrated tasks. One study found that content-related manipulation of listening sources is positively related to speaking proficiency (Frost, Elder, & Wiggleworth, 2011). However, it should be noted that the source material used in the study by Frost et al. was only a single short recording from a radio programme, which may not be sufficient to measure test-takers' listening competence. We are interested here in identifying the cognitive processes involved in listening, how they affect the written product of integrated tasks and to what extent the listening competence of test takers can predict their performance in integrated writing.

In recent years, researchers have also studied the construct validity of integrated tasks by examining discourse features. A study revealed that the written products of both independent and integrated tasks seem to share construct coverage of discourse features, although they "tap into different elements of writing" (Guo, Crossely, & McNamara, 2013, p. 234). Correlation studies have found that lexical, syntactic, rhetorical, and pragmatic features differ significantly between independent and integrated writing tasks (Cumming et al., 2005). Features, such as lexical sophistication and diversity (Guo, Crossely, & McNamara, 2013; Yu, 2013a), as well as fluency, grammatical accuracy, syntactic complexity, and verbatim source use (Gebril & Plakans, 2009, 2013), seem to be good indicators of test-takers' language proficiency level and can predict scores for both independent and integrated tasks.

A related field of studies focused on discourse syntheses involved in the composing process of integrated tasks (Asencion, 2004; Esmaeili, 2002; Plakans, 2008, 2009a, 2009b; Plakans & Gebril, 2012). For instance, by adapting Spivey's (1997) discourse synthesis framework, which was developed from first language research, Asencion (2004), Plakans (2009b), Yang (2009) and Yang and Plakans (2012) revealed that second language learners with higher proficiency levels generally employ the same three subprocesses (i.e., organizing, selecting, and connecting information from source texts) as first language users would do during the composing process. Research on specific task types, such as summarization, has also highlighted discourse synthesis as an important element in the process of composing (van Dijk & Kintsch, 1983; Yu, 2013a). Drawing on previous studies, Knoch and Sitajalabhorn (2013) noted the prominence of skills required to mine sources and select ideas, synthesize ideas, choose the organizational structure, and connect the ideas; thus, they proposed that these features should be included in the test construct.

Existing research in the field of integrated tasks has also shown that the majority have been conducted in the context of English as a second language (ESL). Few have looked at the testing of English as a first language, but almost none has dealt with Chinese Language testing. With the growing trend of learning Chinese worldwide and the preference for integrated writing as a means of assessment, research on integrated writing tasks in Chinese is likely to provide us with further insights into the task construct of integrated writing assessment and complement the previous studies in the field of English as a foreign/second language assessment. One can imagine that both first and second language test takers may present certain similar and different characteristics in the process of completing an integrated task. On one hand, Cook (2010) claimed that the acquisition of first and second languages is perceived to have intrinsic and unavoidable differences. On the other hand, integrated tasks in both first and second languages also have similarities, so the study of the process for the first language can serve as a reference for the second. To quote an example, although Spivey's (1997) research was conducted in a first language context, it shows that the three sub- processes mentioned above are also evident in second language test takers; the latter also faced difficulties in style, vocabulary, and selecting adequate expressions (Plakans, 2009b). Therefore, we are interested in investigating first language test takers because findings from first language research can inform the

teaching, learning, and assessment of the language concerned as a second language, especially for those students with high levels of proficiency.

Research context and questions

In the last century, the Chinese Language curriculum in Hong Kong emphasized reading skills, writing skills, use of reference books and extensive reading, with the aim of cultivating language proficiency (Hong Kong Curriculum Development Committee, 1975; Hong Kong Curriculum Development Council, HKCDC, 1990). Accordingly, the assessment of reading and writing skills has always been the focus of Chinese Language in the Hong Kong Certificate of Education Examination (HKCEE), one of the public examinations in secondary schools. Two separate test papers, Writing and Language Use and Reading Comprehension and Prescribed Text Questions, were used in Chinese Language examinations before 2007. However, in the early 21st century, HKCDC & Hong Kong Examinations and Assessment Authority (HKEAA) (2007) posited that the Chinese Language curriculum should take a much broader and comprehensive view of language competence in its teaching and assessment; as a result, three new papers (listening, speaking, and integrated writing) were added to the assessment of Chinese Language. The integrated writing is designed to assess not only students' abilities in integrating information from different sources but also their higher-order thinking skills. In the integrated writing, test takers first listen to a recording, read several passages (including diagrams) and write an article in the form of a report, speech, or letter to an organization. In 2007, the HKEAA and the Education and Manpower Bureau (EMB) (now the Education Bureau, EDB) commissioned scholars in tertiary institutions to develop assessment standards for Chinese Language competence. Zhu (2005) developed the "standards of Chinese Language assessment in reading, writing and integrated skills," which were implemented in the HKCEE Chinese Language Level Descriptors and Exemplars for Standards-Referenced Assessment (HKEAA, 2005). In 2012, the HKCEE and Hong Kong Advanced Level Examination (HKALE) were incorporated into the new Hong Kong Diploma of Secondary Education (HKDSE) Examination. After some fine-tuning, these standards are still in use in the current HKDSE. Our previous research, a survey with 732 Hong Kong Secondary Four students, showed that the integrated writing paper was regarded as significantly more difficult than the other four papers, regardless of students' achievement level and gender (Zhu & Wu, 2013).

The number of test papers (i.e., five) in one examination has prompted some heated debates among teachers. Some feel that the listening paper (to distinguish it from the listening-reading- writing paper, we call it the Independent Listening Task, hereafter) is the least useful in assessing students' language competence because Chinese is their native language. They doubt if the Independent Listening Task is able to classify students with different language proficiencies and whether it simply repeats the assessment of the construct of listening, which is also included in the Listening-Reading-Writing Task. Some teachers even suggest that the Independent Listening Task should therefore be removed. To understand to what extent the teachers' dissatisfaction is based on sufficient empirical evidence as well as to address the research gap we identified in our literature review, the present study aimed to (a) explore the relationships between the Listening-Reading- Writing Task and the Independent Listening Task and (b) to examine the language competences assessed by the Listening-Reading-Writing Task. Specifically, we ask four research questions:

(a) Is there a meaningful relationship between the scores in the Independent Listening Task and Listening-Reading-Writing Task in the testing of Chinese as a first language?

(b) Are there any common factors in the competence assessed by the Independent Listening Task and the Listening-Reading-Writing Task?

(c) What competence factors are assessed by the Listening-Reading-Writing Task?

(d) Which, if any, performance indicators of the Independent Listening Task can significantly predict scores in the Listening-Reading-Writing Task?

Method

Participants

The participants in the study were 285 Secondary Five students (average age 17) from six Hong Kong secondary schools, two each from Bands 1 to 3. In Hong Kong, secondary schools are classified into three bands based on students' performance in an area-wide academic aptitude test, with Band 1 schools admitting mostly high-ability students and Bands 2 and 3 admitting mostly moderate- and low-ability students, respectively. Among the six selected schools, four (two Band 1 and two Band 3) were located in public housing areas and the two Band 2 schools in private housing areas, implying that students from different socioeconomic backgrounds had been recruited. All schools were government aided, with the exception of one Band 2 school, which was directly subsidized by government. In each of the Band 1 and 3 schools, one Secondary Five class was randomly selected; in each Band 2 school, two classes were randomly selected. Table 1 summarizes demographic information about the participants. Fifty-nine students were excluded from analysis due to incomplete data. Of the remaining 226 students whose responses were complete, 121 were boys and 105 girls.

Table 1. Number of participants from every band of schools.

	Band 1		Band 2		Band 3		
	School 1	School 2	School 3	School 4	School 5	School 6	Total
Female	17	17	33	22	6	10	105
Male	17	18	31	32	6	17	121
Total	34	35	64	54	12	27	226

Instruments

Two tasks, namely, the Independent Listening Task and the Listening-Reading-Writing Task, were developed to assess Hong Kong secondary school students' performance in listening and integrated writing. The tasks are competence-based and aligned with the principles stipulated in the new Chinese Language curriculum in Hong Kong (HKCDC & HKEAA, 2007; HKCDC, 2001a, 2001b).

The Independent Listening Task

Based on the nature of the information stated in the listening source, listening comprehension can be divided into two broad categories: "direct meaning comprehension" refers to comprehending surface information that is explicitly stated in the input text, whereas "inferred meaning comprehension" refers to understanding implicit information that is not so clearly stated (Weir, 1993).

Listening is a process in which listeners actively manipulate linguistic and nonlinguistic knowledge to construct shared mutual beliefs (Brown, 1995; Vandergrift, 1999). For how each type of knowledge is employed by listeners in comprehension, researchers have proposed three models of listening comprehension. The first is the bottom-up model, which assumes that the listening process starts with the lowest level of detail (e.g., acoustic input) and moves up to the highest (e.g., the communicative situation). The second is the top-down model, which assumes that listeners apply their nonlinguistic knowledge to comprehend a text by interpretation, prediction, and hypothesis testing (Alderson, 2000; Buck, 2001). The third is the interactive model, which proposes that the cognitive actions involved in listening can take place in any order, simultaneously or cyclically (Grabe, 1991). Buck's (2001) interpretation of the listening comprehension process is widely accepted:

> To summarize the process, the listener takes the incoming data, the acoustic signal, and interprets that, using a wide variety of information and knowledge, for a particular communicative purpose; it is an

> inferential process, an ongoing process of constructing and modifying an interpretation of what the text is about, based on whatever information seems relevant at the time. (p. 29)

Listening is usually perceived as a very simple modality in first language learning; thus, little focus has been placed on studying learners' listening competence and their cognitive processes during listening assessment tasks. Both listening and reading comprehension involve a similar process of taking incoming data and interpreting them. Listening is unique because, unlike reading, a number of factors affect comprehension. First, the listener has little or no control over the input speed of speaking material. Second, they often listen only once and, therefore, cannot pause to work out the meaning of the material, as one can when reading. In a listening test, test takers undergo a cognitive process of constructing and modifying their interpretation of the sources within a limited time. This can be a complicated process, especially when working out answers that require higher-order thinking skills, such as inferring meaning.

In the educational objectives of cognitive domain, Bloom's taxonomy comprises knowledge, comprehension, application, analysis, synthesis, and evaluation (Bloom, 1956), and its revised version includes remembering, understanding, applying, analysing, evaluating, and creating (Anderson & Krathwohl, 2001). Both schemes highlight the higher-order thinking skills of learners. Drawing on research on listening process and the educational objectives of the cognitive domain, we proposed a six-level listening comprehension competence framework entitled the "Six Types of Listening Comprehension Processes" (Zhu, 2012). We defined listening competence in the following terms: (a) memorization (retaining and retelling explicit information with phrases and sentences from the listening texts); (b) explanation (paraphrasing important phrases and sentences in one's own words); (c) summarization (summarizing the theme of the text and sorting out the ideas and the interrelationship of the content); (d) elaboration (inferring the implied meanings and purposes by applying imagination and inference to the surface meaning in listening material); (e) evaluation (appreciating and criticising the views and attitude of the speaker and his/her language use); and

(f) creation (generating personal opinions, or solving real-life problems by applying the information provided). According to the complexity of the thinking process involved, we argued that the first two are related to lower-order thinking skills, while the other four are related to higher-order thinking skills. According to the depth of listening comprehension, we argued that the first three processes are prerequisites for students to obtain basic and textual comprehension of the listening input, while the other three are needed to move beyond textual comprehension.

An Independent Listening Task in the present study was developed to assess students' listening competence, based on the Six Types of Listening Comprehension Processes. The participants were asked to respond to 14 items after listening to two recordings on the same topic. The number of items for each indicator (or type) of listening competence was as follows; memorization (2 items; 1 of which comprises 4 sub-items), explanation (2), summarization (3), elaboration (4), evaluation (2), and creation (1). Among these items, nine were multiple-choice and five were short-answer ques- tions. For the latter, a detailed marking scheme with student answer exemplars was provided. The duration of the Independent Listening Task was 30 minutes. The scores assigned to the indicators memorization, explanation, summarization, and elaboration ranged from 0 to 6, whereas evaluation and creation ranged from 0 to 8. Cronbach's alpha for the 14-item task was a moderate .68, which is within the acceptable range (DeVellis, 2003).

The Listening-Reading-Writing Task

The design of the Listening-Reading-Writing Task was aligned with the integrated writing paper of Chinese Language in the HKDSE (HKEAA, 2012). It assesses students' competence in integrated writing, based on their comprehension and use of listening and reading source materials. The 60-minute task required test takers to listen to an audio recording of a dialogue among students from a

Hong Kong secondary school. Among these students, Cheung Zit Yin and Li Su Ching disagreed on which landscape to preserve in the planned campus refurbishment of the school. Cheung Zit Yin supported the preservation of ancient trees, whereas Li Su Ching supported the preservation of a historical lotus pond. After the listening, test takers then read five texts. Text 1 was a notice in the school's student magazine inviting submission of articles on views about the school's refurbishment project. Text 2 presented two contrasting viewpoints on the refurbishment work: one from the principal and the other from an alumnus, each arguing through their propositions. Text 3 was about the school's brief history. Text 4 illustrated campus landscapes with several photographs and descriptions. Text 5 was the programme of the school's Open Day. The five texts comprised approximately 2,400 Chinese characters. Test takers were requested to write an article of 400 words or more in the persona of Cheung Zit Yin or Li Su Ching to express their views on which types of historic campus landscape to be preserved.

When developing the above-mentioned "standards of Chinese Language assessment in reading, writing and integrated skills", Zhu (2005) proposed his four traits of integrated writing competence for secondary school students, or "four pillars"(四条柱, si tiao zhu) as widely known among the Chinese Language teaching and assessment sectors in Hong Kong. These four pillars are contextual awareness, citation and synthesis, original opinion and argument, and written expression and organization. In the present study, we adopted these four pillars with 10 performance indicators that were used to assess students' performance on the integrated writing task. Details of the indicators are shown in Table 2. The scores assigned to each of the 10 indicators ranged from 0 to 10.

Table 2. Ten indicators of the Listening-Reading-Writing Task.

Indicators	Description
Identification (W1)	To be aware of writer and reader
Tone (W2)	To apply an appropriate tone that is consistent with the identity of the writer and his relationship with the readers
Writing convention (W3)	To demonstrate standard usage and mechanics of Chinese Language practical writing
Interaction (W4)	To address the readers' concerns, in relation to the context and purpose of writing
Synthesis (W5)	To extract and summarize ideas and/or information relevant to the topic
Citation (W6)	To select relevant ideas and/or information, including lifting key words from the sources or using their own words for expression; connecting the ideas to his own experience
Original opinion (W7)	To infer and predict by making reference to the sources; make thoughtful, practical and creative suggestion
Argument (W8)	To provide concrete evidence and convincing explanation
Language use (W9)	To write accurately, concisely and fluently
Organization (W10	To cohesively present ideas, with a structure that clearly connects the main ideas and other details in an orderly manner

Piloting the instruments

A pilot study was conducted with 36 students in two Secondary Five classes in a Band 2 school. Both the tasks and the marking scheme of the independent listening task were adjusted according to the students' performance in the pilot study. For the Independent Listening Task, the amendments made included (a) standardizing the expression used in the options of one of the multiple-choice questions, (b) converting one short-answer question to a fill-in-the-blank question to avoid the ambiguity, and (3) specifying the requirements of a short-answer question.

Scoring procedures

We held focus-group discussions with five raters to ensure they understand and implement the marking schemes consistently. All the raters had more than six years of teaching experience in secondary schools. At the focus group, we discussed the rationales for the overall design of the Independent Listening Task and the Listening-Reading-Writing Task, the selection of the listening and reading texts, the difficulty level of the tasks and the development of the marking schemes. Typical students' answer scripts were included as exemplars in the marking scheme of the Independent Listening Task and the rubric of the Listening-Reading-Writing Task for raters' reference during marking.

Two raters double marked 10 sample Independent Listening Task scripts. The other three raters each marked 10 Listening-Reading-Writing Task scripts. The researchers and raters then reviewed the results, discussed any discrepancies between the raters, and arrived at agreed scores for the scripts. All 226 Independent Listening Task and Listening-Reading-Writing Task scripts were randomly allocated to raters, with each rater assigned an approximately equal number of participants from schools in Bands 1, 2, and 3. In addition, 5% of scripts were randomly selected and marked by all raters to monitor consistency. Cronbach's alpha for the Listening-Reading-Writing Task was .82.

Data analysis

The data were coded and entered into SPSS for statistical analysis. Descriptive statistics (mean, standard deviation, skewness, and kurtosis) were first calculated to examine the central tendencies, variation, and distributional properties of the data. We then conducted Pearson product-moment correlation analysis to examine the association between the scores in the Independent Listening Task and the Listening-Reading-Writing Task. If no significant correlation was found, this would imply the possibility that the Independent Listening Task and the Listening-Reading-Writing Task tap

Table 3. Research questions and the respective statistical tests used.

Research questions	Statistical tests
Is there a meaningful relationship between the scores in the Independent Listening Task and Listening-Reading-Writing Task in the testing of Chinese as a first language?	Descriptive statistics; Pearson product-moment correlation analysis
Are there any common factors in the competence components assessed by the Independent Listening Task and Listening- Reading-Writing Task?	Joint factor analysis with all the indicators from the Independent Listening Task and the Listening-Reading- Writing Task
What competence factors are assessed by the Listening- Reading-Writing Task?	Exploratory factor analysis of the Listening-Reading-Writing Task
Which, if any, performance indicators of the Independent Listening Task can significantly predict scores in the Listening-Reading-Writing Task?	Regression analysis of total score of the Listening-Reading- Writing Task on the indicators of the Independent Listening Task

different constructs; otherwise, a joint factor analysis would be conducted. Joint factor analysis refers to a technique in which items from several measures are entered simultaneously into a factor analysis to determine their factor structure across the measures. This approach has been used extensively in research related to intelligence (see, e.g., Kaufman, Ishikuma, & Kaufman, 1994; McGhee, 1993), personality (see, e.g., Ferguson, 2001; Taylor, 1996), and anxiety and depression (see, e.g., Stark & Laurent, 2001) to assess structural similarities between different instruments seeking to measure the same or similar constructs. For instance, in personality research, the well-known Big Five model of personality (extraversion, agreeableness, conscientiousness, neuroticism, and openness) came from

the application of joint factor analysis with different taxonomies of personality used in different instruments (McCrae, 1989). In the present study, joint factor analysis was conducted to investigate the emerging structure of the Independent Listening task and the Listening-Reading-Writing Task. All indicators from both tasks were entered simultaneously into the factor analysis to determine whether common underlying factors could be identified across them. A separate factor analysis of the Listening-Reading-Writing Task was then performed to examine its own structure. Finally, a regression analysis of test-takers' total Listening-Reading-Writing Task score on the indicators of the Independent Listening Task was conducted for further exploration of the relationship between the two tasks. Table 3 links the research questions to the statistical tests performed in the present study.

Results

Descriptive statistics

The descriptive statistics for all indicators of the Independent Listening Task and the Listening-Reading-Writing Task are presented in Table 4. The percentage scores on memorization (L1), explanation (L2), and summarization (L3) in the Independent Listening Task were relatively high (more than 80.0%). These three indicators involved the cognitive processes of memorizing, comprehending, and summarizing, respectively; in other words, they were indicators of the students' performance on basic and textual comprehension of the audio recordings. The other three indica- tors, elaboration (L4), evaluation (L5), and creation (L6), the cognitive processes of inferring, evaluating, and creating (Zhu, 2012), had relatively low percentage scores. In the Listen-Reading- Writing Task, identification (W1) received 83.0%, the highest percentage score. The Original opinion (W7) and Argument (W8), which involved the cognitive processes of creating and evaluating showed relatively low percentage scores of 32.5% and 30.3%, respectively. The percentage scores of the remaining indicators ranged from 49.3% to 56.5%. In both tasks, standard deviations ranged from

1.02 for language use (W9) to 3.18 for identification (W1). All absolute values of skewness and kurtosis were less than 2.00, which are within the accepted range for univariate normality (Byrne, 1998; Kline, 2005).

Table 4. Descriptive statistics for indicators all .

	Full score	Mean (convert to percentage score)	Std Dev	Skewness	Kurtosis
Independent Listening Task					
Memorization (L1)	6	4.94 (82.3)	1.18	−1.09	.73
Explanation (L2)	6	4.96 (82.7)	1.55	−1.31	.74
Summarization (L3)	6	5.24 (87.3)	1.19	−1.44	1.68
Elaboration (L4)	6	3.90 (65.0)	1.05	−.43	−.06
Evaluation (L5)	8	3.92 (49.0)	1.59	.01	−.23
Creation (L6)	8	4.57 (57.1)	2.64	−.28	−.80
Listening-Reading-Writing Task					
Identification (W1)	10	8.30 (83.0)	3.18	−1.67	1.47
Tone (W2)	10	5.42 (54.2)	1.35	.12	1.64
Writing convention (W3)	10	5.65 (56.5)	2.30	−.56	−.50
Interaction (W4)	10	4.93 (49.3)	1.57	−.07	−.05
Synthesis (W5)	10	5.18 (51.8)	1.73	−.19	−.11
Citation (W6)	10	4.95 (49.5)	1.81	−.21	−.02
Original opinion (W7)	10	3.25 (32.5)	2.05	−.17	−1.04
Argument (W8)	10	3.03 (30.3)	1.99	−.04	−1.02
Language use (W9)	10	5.60 (56.0)	1.02	−.29	1.74
Organization (W10)	10	5.56 (55.6)	1.34	−.67	.85

Note. Percentage score = Mean/Full score *100.

Pearson product-moment correlation

A Pearson product-moment correlation analysis was performed to examine the relationships between the Independent Listening Task and the Listening-Reading-Writing Task. As shown in Table 5, among the first three indicators of the Independent Listening Task (memorization, L1; explanation,

L2; and summarization, L3), only explanation (L2) had a significant correlation with one indicator of the Listening-Reading-Writing Task, that is, interaction (W4). We argued earlier that these three indicators are processes for students to obtain basic and textual comprehension of the listening input. However, the other three indicators of the Independent Listening Task, elaboration (L4), evaluation (L5), and creation (L6), which process beyond textual comprehension of the listening input, were all significantly correlated with multiple indicators of the Listening-Reading-Writing Task. Specifically, both elaboration (L4) and creation (L6) were significantly correlated with five indicators of the Listening-Reading-Writing Task, with elaboration (L4) correlated significantly with tone (W2), interaction (W4), synthesis (W5), citation (W6), and language use (W9); and creation (L6) with tone (W2), original opinion (W7), argument (W8), language use (W9), and organization (W10). Furthermore, evaluation (L5) was significantly correlated with all the 10 indicators but identification (W1) of the

Table 5. Pearson product-moment correlations between the Independent Listening Task and the Listening-Reading-Writing Task (N = 226).

	Memorization (L1)	Explanation (L2)	Summarization (L3)	Elaboration (L4)	Evaluation (L5)	Creation (L6)	Total
Identification (W1)	.00	.04	−.07	.10	.07	.08	.10
Tone (W2)	.11	.13	.13	.26***	.19**	.16*	.29**
Writing convention (W3)	.05	.03	−.11	.00	.15*	.01	.06
Interaction (W4)	.10	.15*	.10	.17*	.22***	.12	.25**
Synthesis (W5)	.04	.05	.06	.15*	.21***	.13	.22**
Citation (W6)	.07	.11	.09	.17**	.25***	.11	.25**
Original opinion (W7)	.06	.05	.07	.04	.14*	.15*	.18**
Argument (W8)	.10	.06	.05	.07	.15*	.18**	.22**
Language use (W9)	.11	.03	.09	.20**	.22***	.15*	.26**
Organization (W10)	.00	.03	.05	.11	.24***	.16*	.23**
Total	.09	.10	.04	.17**	.26**	.18**	.29**

* p < .05, ** p < .01, *** p < .001.

Listening-Reading-Writing Task. However, the significant correlations were all weak, ranging from .14 between evaluation (L5) and original opinion (W7) to .26 between elaboration (L4) and tone (W2). In the total scores achieved by the students in each of the two tasks, we noticed that the Independent Listening Task and the Listening-Reading-Writing Task were significantly correlated (r = .29, p < .01) with a small magnitude. To further examine if it was sufficient to enable the extracting of common factors between the two tasks, we ran a follow-up joint factor analysis.

Joint factor analysis of both tasks

A Kaiser-Meyer-Olkin (KMO) coefficient greater than .5 indicates that the variables being analysed belong together psychometrically (Kaiser, 1970, 1974). A significant result for the Bartlett test of sphericity indicates that the variables of interest within the sample matrix are not independent (Bartlett, 1954). In this present study the indicators of the Independent Listening Task and the Listening-Reading-Writing Task had a KMO measure of sampling adequacy of .75 and a significant Bartlett test of sphericity ($\chi 2$ = 1706.33, df = 120, p < .001), suggesting that the data would be appropriate for factor analysis. Joint factor analysis would help to clarify the relationships among these groups of indicators. Factor analysis with direct oblimin for rotation was first run; it was found that the majority of interfactor correlations (see Table 6) were low, with one exception, which was more than .50. Overall, it suggests that the use of orthogonal varimax rotation with Kaiser Normalization would be more appropriate (Hair, Black, Babin, & Anderson, 2014; Tabachnick & Fiddell, 2013).

The results of the joint factor analysis showed that there were five eigenvalues greater than 1; they were 4.72, 1.72, 1.55, 1.14, and 1.08. As shown in Table 7, the five-factor solution seemed to explain well the "four pillars" of the integrated writing competence as identified in our previous research (Zhu, 2005), and the additional component of the Independent Listening Task. These five factors together, accounted for 63.8% of the variance.

The indicators of the Independent Listening Task loaded on factor 5, which we call the listening competence factor. It seemed to be quite distinct from the other four factors, which are all related to the Listening-Reading-Writing Task. The listening competence factor loadings of the indicators of the Independent Listening Task (i.e., L1 to L5) ranged from .33 to .60, with no substantial crossloading on other factors. The indicator creation (L6) had a low loading (.20) on factor 5.

The Listening-Reading-Writing Task was represented by four underlying factors. The first factor was primarily associated with identification (W1) and writing convention (W3), which we call it contextual awareness. The second factor was associated with synthesis (W5) and citation (W6), which we call citation and synthesis. The third factor was mainly associated with original opinion (W7) and argument (W8), we call original opinion and argument. The fourth factor was associated with tone (W2), interaction (W4), language use (W9), and organization (W10), we call written expression and organization. These four factors corresponded well with the notion of "four pillars" underlying the integrated writing task. All the 10 indicators of Listening-Reading-Writing Task had relatively large loadings on their respective primary factors, ranging from .49 (identification, W1, on its primary factor of contextual awareness) to .96 (argument, W8, on its primary factor of original opinion and

Table 6. Interfactor correlation matrix of the Independent Listening Task and the Listening-Reading-Writing Task (N = 226).

Factor	1	2	3	4	5
1	1.00				
2	.21	1.00			
3	−.10	−.27	1.00		
4	.22	.52	−.34	1.00	
5	−.08	.17	−.18	.35	1.00

Note. Extraction method: principal axis factoring. Rotation method: Direct Oblimin with Kaiser Normalization.

Table 7. Joint factor analysis of the Independent Listening Task and the Listening-Reading-Writing Task (N = 226).

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Independent Listening Task					
Memorization (L1)	.02	−.01	.04	.03	.33
Explanation (L2)	.06	.01	.01	.00	.45
Summarization (L3)	−.20	.01	.03	.09	.39
Elaboration (L4)	.00	.08	−.04	.12	.60
Evaluation (L5)	.05	.12	.08	.19	.34
Creation (L6)	−.02	.06	.13	.14	.20
Listening-Reading-Writing Task					
Identification (W1)	.49	.39	−.02	.22	−.01
Tone (W2)	.14	.14	−.03	.78	.24
Writing convention (W3)	.71	.02	.15	.20	.01
Interaction (W4)	.15	.11	.16	.51	.23
Synthesis (W5)	.10	.93	.18	.26	.09
Citation (W6)	.07	.83	.16	.34	.16
Original opinion (W7)	.08	.11	.93	.15	.07
Argument (W8)	.10	.16	.96	.17	.10
Language use (W9)	.04	.26	.14	.76	.12
Organization (W10)	.22	.26	.25	.70	.02

Note. Factor 1 = *contextual awareness*, Factor 2 = *citation and synthesis*, Factor 3 = *original opinion and argument*, Factor 4 = *written expression and organization*, Factor 5 = *listening competence*.

argument). However, two indicators, identification (W1) and citation (W6), which showed cross-loadings greater than .30 on two factors, are worth noting. Specifically, identification (W1) cross-loaded on contextual awareness (0.49) and citation and synthesis (.39). Citation (W6) crossloaded on citation and synthesis (0.83) and written expression and organization (.34). We argue that the

crossloading of identification (W1) on the two factors was probably due to test-takers' awareness and consideration of their relationship with their readers because their awareness and consideration of their audience could affect the kind of information they would cite and synthesize in their writing. As for the second crossloading, we anticipate that citation (W6) and processing of a particular type of information could affect how students organize and present this information (by copying or imitating, probably) in their writing

To summarize, our data suggested that all the indicators of the Independent Listening Task were on the fifth single factor (listening competence), while those indicators of the Listening-Reading- Writing Task formed the first four factors with two meaningful crossloadings. The listening competence factor had no commonality with the integrated writing task.

Separate factor analysis of integrated writing task

Because the Listening-Reading-Writing Task shared no common factor with the Independent Listening Task as reported above, the question then arose as to what the Listening-Reading- Writing Task had actually assessed. To answer this question, a separate factor analysis was conducted to explore its underlying structure. Similar to the joint factor analysis, we also tried direct oblimin rotation method first. The interfactor correlation matrix (see Table 8) showed that the correlations were low; therefore, we ran orthogonal varimax rotation with Kaiser Normalization (KMO = .76; Bartlett test of sphericity, $\chi 2 = 1553.72$, df = 45, $p < .001$).

As expected, a four-factor model emerged (eigenvalues ranging from 4.45 to 1.06), accounting for 81.8% of the variance. The factors emerging from the indicators of the Listening-Reading-Writing Task in the separate factor analysis were similar to those in the joint factor analysis (see Table 8). As shown in Table 9, there were three indicators of integrated writing task that had crossloadings in relation to W1, W6 and W10. The loadings for the primary factor were about twice the loadings for the secondary factor.

Table 8. Interfactor correlation matrix of the Listening-Reading-Writing Task (*N* = 226).

Factor	1	2	3	4
1	1.00			
2	.34	1.00		
3	−.44	−.24	1.00	
4	.43	.24	−.31	1.00

Note. Extraction method: principal axis factoring. Rotation method: Direct Oblimin with Kaiser Normalization.

Table 9. Factor analysis of the Listening-Reading-Writing Task (*N* = 226).

	Factor 1	Factor 2	Factor 3	Factor 4
Identification (W1)	.77	.32	−.03	.13
Tone (W2)	.15	.15	−.02	.84
Writing convention (W3)	.49	−.01	.17	.22
Interaction (W4)	.11	.13	.17	.56
Synthesis (W5)	.18	.90	.17	.25
Citation (W6)	.13	.86	.16	.34
Original opinion (W7)	.09	.12	.93	.14
Argument (W8)	.10	.17	.96	.17
Language use (W9)	.13	.28	.14	.72
Organization (W10)	.32	.24	.25	.64

Note. Factor 1 = *contextual awareness*, Factor 2 = *citation and synthesis*, Factor 3 = *original opinion and argument*, Factor 4 = *written expression and organization.*

Regression analysis

The Pearson product-moment correlation analysis established that there was a connection between the Independent Listening Task and Listening-Reading-Writing Task, although the association between

the two tasks was not sufficient to generate a common factor by employing the joint factor analysis. The next research question asked whether test-takers' performance on the Independent Listening Task could significantly predict participants' performance on the Listening-Reading-Writing Task. As we reported earlier, elaboration (L4), evaluation (L5), and creation (L6) of the Independent Listening Task performance were significantly correlated with multiple indicators of the Listening-Reading-Writing Task. After checking that the data met the assumptions of regression analysis (i.e., linearity, homoscedasticity, and normality), a regression analysis of the Listening-Reading-Writing Task score was performed on the scores of the three indicators of the Independent Listening Task.

As shown in Table 10, elaboration (L4), evaluation (L5), and creation (L6) explained 10% of the variance of the scores in the Listening-Reading-Writing Task, indicating low level of predictive power. Both evaluation (L5) ($\beta = .21$, $t = 3.20$, $p < .01$) and creation (L6) ($\beta = .14$, $t = 2.13$, $p < .05$) significantly predicted test-takers' scores in the Listening-Reading-Writing Task, accounting for 8.9% variance of the integrated writing task. However, elaboration (L4) did not have significant predict-ability. This indicates that if test takers took both the Independent Listening Task and Listening-Reading-Writing Task, their performance in evaluation (L5) and creation (L6) tended to be

Table 10. Regression analysis of total score of the Listening-Reading-Writing Task on three indicators of the Independent Listening Task ($N = 226$).

	B (unstandardized regression coefficient)	Standard error	β (Standardized regression coefficient)	t	
Elaboration (L4)	1.11	.75	.10	1.48	
Evaluation (L5)	1.59	.50	.21**	3.20	
Creation (L6)	.62	.29	.14*	2.13	
R^2					.10
F					8.07

Note. * $p < .05$, ** $p < .01$.

consistent with their performance on the Listening-Reading-Writing Task overall. These results may suggest that while test takers performed well in the Independent Listening Task, it does not always imply that they will do well in the Listening-Reading-Writing Task.

Discussion and conclusion

The general relationship between Independent Listening Task and Listening-Reading-Writing Task

A statistically significant but small correlation ($r = .29$, $p < .01$) was found between the total scores on the Independent Listening Task and the Listening-Reading-Writing Task, indicating there does not seem to be strong evidence supporting that the two tasks are repetitive. Looking into the details, indicators at the basic and textual comprehension level, that is, memorization (L1), explanation (L2), and summarization (L3) of the Independent Listening Task, do not have meaningful associations with any of the indicators of the Listening-Reading-Writing Task, with the only exception of a significant relationship between explanation (L2) and interaction (W4). However, beyond the textual comprehension level, all other three indicators of the Independent Listening Task, that is, elaboration (L4), evaluation (L5), and creation (L6), correlates significantly with a number of indicators of the Listening-Reading-Writing Task. This seems to imply that, compared to basic and textual comprehension, the listening skills required beyond textual comprehension have a stronger positive effect on completing the Listening-Reading-Writing task.

Particularly, we should further note the following: (a) The significant, but small magnitude of correlations between the three indicators elaboration (L4), evaluation (L5), and creation (L6) of the listening task and multiple indicators of the integrated writing task indicate that the requirement and demand of comprehension beyond basic understanding of the listening input in the Independent Listening Task and the Listening-Reading-Writing Task were not the same. (b) Being able to evaluate text (evaluation, L5) and coming up with original ideas (creation, L6) as measured in the Independent

Listening Task were able to predict significantly test-takers' performance on the Listening-Reading-Writing Task, with the two indicators accounting for 8.9% of the variance of the total score on the integrated writing task. However, elaboration (L4) did not significantly predict test-takers' total score in the Listening-Reading-Writing Task. The reason may be that the listening input in the Listening-Reading-Writing Task tended to provide explicit information and ideas for test takers to cite or synthesize (Zhu, 2005); thus, there may not be a great need to infer meanings from the listening sources. On the basis of the above overall results, we may suggest that higher- order thinking skills are more important for successful performance in the Listening-Reading- Writing Task than lower-order thinking skills, because test takers are required to evaluate the views in the sources and make new ideas based on the source.

The joint factor analysis showed that the indicators of the Independent Listening Task formed one factor, and the indicators of the Listening-Reading-Writing Task four factors. This finding indicates that there was little overlapping in what the Independent Listening Task and Listening- Reading-Writing Task assess. It also suggests that the integrated tasks in a first language (in this case, Chinese) may require a different threshold level from similar tasks in a second language. Cumming (2013) pointed out that one of the perils of integrated writing assessment was that test takers have to reach a certain threshold level before they can perform the integrated writing tasks. In other words, if a second language learner does not have the sufficient listening comprehension ability, his or her performance in the integrated listening-reading-writing tasks that require listening ability could be hampered. In our case, however, the test takers have already acquired sufficient listening ability to understand the input that was delivered in their first language—Chinese. To some extent, the potential difficulty that they might face in achieving basic understanding of the listening input in their first language could be quite different from the difficulty that second language learners might have to face.

The construct assessed by the integrated writing task

The joint factor analysis on the data from the Independent Listening Task and the Listening- Reading-Writing Task, as well as the separate factor analysis on the data from the Listening- Reading-Writing Task, provided empirical evidence supporting the existence of the "four pillars" in integrated writing competence, that is, contextual awareness, citation and synthesis, original opinion and argument, and written expression and organization (Zhu, 2005). The Listening- Reading-Writing Task does not assess isolatedly language skills. Instead, it assesses the integrated language competence of test takers; that is, to evaluate test-takers' comprehensive employment of multiple language skills to complete authentic tasks (HKEAA, 2012).

The contextual awareness factor was found in the task requirement that asks the students to write a practical article (e.g., a report or speech) with an awareness of themselves being the author of the article and its readership. Test takers were required to consider the appropriateness of their use of language while communicating with their readers through the article. Because Chinese culture emphasizes personal, social, and governmental morality, and test takers are expected to show their respect in their writing toward people with seniority and also care for the young. As Yu (2013b) suggested, "What roles do test takers' characteristics (e.g., language and scientific skills, social, educational background, and training experience) play in their performance?" (p.113) is an important factor to consider when defining and operationalizing the construct of integrated writing assessment. We argue that contextual and cultural awareness of readership may be even more important for integrated writing tasks in a first language than a second language because it is likely that first language learners have more resources (including their language proficiency in the first language) at their disposal and therefore may pay more attention to cultural and contextual factors when they organize and synthesize the source content. However, in second language assessment, as some test takers did not meet the threshold required to successfully complete an integrated writing task

(Cumming, 2013), they have fewer resources to attend to the cultural and contextual factors in their writing.

The citation and synthesis factor refers to an assessment of the comprehension and use of input materials. Our data showed that this factor makes a large contribution to test performance in the integrated writing task. Citation and synthesis in this study is similar to the processes of selecting and connecting in the discourse synthesis framework proposed by Plakans (2009b). Investigating how source text is used in an integrated writing task, Plakans and Gebril (2013) showed that the feature of using listening and reading texts explains over 50% of the variance in scores on reading-listening-writing tasks, and most of the variance is explained by the use of the listening text and the inclusion of important ideas from the sources. In the survey by Zhu and Wu (2014), test takers asserted that they had cited and synthesized from the aural and written sources, but found this process time- consuming and they were uncertain how to perform well.

The original opinion and argument factor captures higher-order thinking skills, such as evaluating and creating. The Listening-Reading-Writing Task requires test takers to present original and creative ideas in their writing rather than simply summarizing the sources. First language learners usually have sufficient listening and reading proficiency and can understand aural and written sources and apply language conventions with ease. They thus have the capacity to pay attention to original opinion and argument in the integrated writing. Originality of thought, development of ideas, and the soundness of the writer's logic are heavily emphasized in first language writing instruction compared to teaching of L2 (Weigle, 2002). In Hong Kong, the integrated task was introduced with the aim of evaluating test-takers' ability, including higher-order thinking skills, to work on authentic language tasks. Correspondingly, the teaching and learning of integrated writing in Hong Kong has placed more emphasis on enhancing synthesizing, creating, and arguing in writing. Integrated tasks in a second language may serve a slightly different purpose. By providing test takers with sources, including content and language, rather than simply giving them a topic to write on, a second language integrated writing task reduces the demand for creativity (Plakans, 2008; Read, 1990). Second language learners tend to copy whole chunks from the sources, and many low- proficiency students plagiarize their sources (Read, 1990). Thus, original opinion and argument is often not evident in second language integrated writing tasks in previous studies.

For the two factors of citation and synthesis and original opinion and argument, teachers tend to recognize that these are the ones which distinguish high from low level of language proficiency. They are also perceived as the two most difficult parts of integrated tasks for students to learn and for teachers to teach and assess (Zhu, 2015; Zhu & Wu, 2014). In the spirit of Assessment for Learning (Black & Wiliam, 2009; Zhu, 2014), while acknowledging the challenges, we propose that these two factors should be emphasized when we design programmes to promote teachers' professional development and students' approaches to learning effectively.

Besides the focus on lexical and conventional features that teachers have always placed emphasis on, the last factor written expression and organization, similar to that in Plakans's (2009b) framework related to organizing, monitoring, and writing, could focus on constructing macrostructure of writing, taking appropriate and consistent tone, presenting ideas cohesively and addressing the readers' concerns to collectively assess students' overall writing abilities (Zhu & Wu, 2014).

The independence of the Independent Listening Task and Listening-Reading-Writing Task in assessing Chinese as a first language

As mentioned in the above correlation analysis, some indicators of the Independent Listening Task and the Listening-Reading-Writing Task were significantly correlated to each other (but with low magnitudes). The follow-up regression analysis revealed that two indicators in the Independent Listening Task, namely, evaluation (L5) and creation (L6), could significantly predict scores in the

Listening-Reading-Writing Task (8.9%). This result has an implication in the context of learning a first language: the Independent Listening Task may reflect higher-order thinking skills, such as evaluation and creation. It also indicates to us that more evidence could be collected to determine whether the factors evaluation (L5) and creation (L6) could be considered to remove from the Independent Listening Task. This would bring out the distinctive assessment objectives of the two tasks and potentially ease the testing burden in practice. Because of the small correlation coefficients between the Independent Listening Task and the Listening-Reading-Writing Task and the low predictability of the Independent Listening Task to the Listening-Reading-Writing Task, we argue that it would be considerably more useful to use both tasks to serve the collective purpose of evaluating language curriculum objectives that comprise listening, speaking, reading, and writing (including integrated writing).

While studying any integrated assessment, we should also look at how the assessment on integrated skills also impacts language teaching and learning (Yu, 2013b). The teaching of reading in Hong Kong has been heavily dictated by commercially produced textbooks and guides, and most of the curriculum time is used to decode the words in passages (Tse, 2009). The present study supports the implementation of the current Chinese Language curriculum and assessment in Hong Kong, which advocates the development of the competences of processing language, four indepen- dent language skills, integrated language and higher-order thinking (e.g., synthesizing, inferring, evaluating, and creating). Other studies (Zhu, 2013, 2015; Zhu & Wu, 2014) have shown that teachers in Hong Kong hold the integrated test in high regard, but face difficulties in fostering the relevant skills and also in designing appropriate tasks in their day to day teaching. Teachers have requested professional support in developing their strategies and skills for teaching integrated writing, and in designing integrated tasks and scoring schemes. In addition, teacher professional development programmes should include information on how learners/test takers would use cog- nitive and metacognitive strategies (Oxford, 1990; Purpura, 1997, 2013) to cope with the demand of integrated assessment tasks, the relationships between cognitive strategies and test performance (e.g., Barkaoui et al., 2013; Purpura, 1998, 1999; van Gelderen et al., 2004; Yang & Plakans, 2012).

We are aware of the limitations of our study. One limitation is related to the small number of indicators within each factor in the integrated writing task. According to Brown (2006), factors that are based on only a few indicators may have determinacy problems and are therefore not stable enough to be replicated across different samples. The small number of participants as well as the small number of test tasks also limit the generalizability of the findings of this study. To achieve a more reliable outcome, further work could be carried out using more indicators of task performance at various levels of task difficulty and complexity, with a larger sample size.

References

Alderson, J. C. (2000). *Assessing reading*. New York, NY: Cambridge University Press.

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching and assessing: A revision of Bloom's taxonomy of educational objectives*. New York, NY: Longman.

Asención, Y. (2004). *Validation of reading-to-write assessment tasks performed by second language learners*

(Unpublished doctoral dissertation). Northern Arizona University, Flagstaff, AZ.

Barkaoui, K., Brooks, L., Swain, M., & Lapkin, S. (2013). Test-takers' strategic behaviors in independent and integrated speaking tasks. *Applied Linguistics*, *34*(3), 304–324. doi:10.1093/applin/ams046

Bartlett, M. S. (1954). A note on the multiplying factors for various X^2 approximations in factor analysis. *Journal of the Royal Statistical Society*, *16*(2), 296–298.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability (formerly: Journal of Personnel Evaluation in Education)*, *21*(1), 5–31. doi:10.1007/s11092-008- 9068-5

Bloom, B. (1956). *Taxonomy of educational objectives (Handbook I: The cognitive domain)*. New York, NY: David McKay.

Brown, A., Iwashita, N., & McNamara, T. (2005). *An examination of rater orientations and test-taker performance on English-for-Academic-Purposes speaking tasks*. Princeton, NJ: Educational Testing Service.

Brown, G. (1995). *Speakers, listeners and communication*. New York, NY: Cambridge University Press. Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press. Buck, G. (2001). *Assessing listening*. New York, NY: Cambridge University Press.

Byrne, B. M. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, application, and programming*. Mahwah, NJ: Lawrence Erlbaum Associates.

Cho, Y. (2003). Assessing writing: Are we bound by only one method? *Assessing Writing*, *8*(3), 165–191. doi:10.1016/ S1075-2935(03)00018-7

Cook, V. J. (2010). The relationship between first and second language acquisition revisited. In E. Macaro (Ed.), *The continuum companion to second language acquisition* (pp. 137–157). London, UK: Continuum.

Cumming, A. (2013). Assessing integrated writing tasks for academic purposes: Promises and perils. *Language Assessment Quarterly*, *10*(1), 1–8. doi:10.1080/15434303.2011.622016

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. doi:10.1016/ j.asw.2005.02.001

DeVellis, R. F. (2003). *Scale development: Theory and applications*. Thousand Oaks, CA: Sage Publications.

Esmaeili, H. (2002). Integrated reading and writing tasks and ESL students' reading and writing performance in an English language test. *Canadian Modern Language Review*, *58*(4), 599–622. doi:10.3138/cmlr.58.4.599

Ferguson, E. (2001). Personality and coping traits: A joint factor analysis. *British Journal of Health Psychology*, *6*(4), 311–325. doi:10.1348/135910701169232

Frost, K., Elder, C., & Wigglesworth, G. (2011). Investigating the validity of an integrated listening-speaking task: A discourse-based analysis of test-takers' oral performances. *Language Testing*, *29*(3), 345–369. doi:10.1177/ 0265532211424479

Gebril, A. (2006). *Independent and integrated academic writing tasks: A study in generalizability and test method*

(Unpublished doctoral dissertation). The University of Iowa, Iowa City, IA.

Gebril, A. (2009). Score generalizability of academic writing tasks: Does one test method fit it all? *Language Testing, 26*

(4), 507–531. doi:10.1177/0265532209340188

Gebril, A. (2010). Bringing reading-to-write and writing-only assessment tasks together: A generalizability analysis.

Assessing Writing, *15*(2), 100–117. doi:10.1016/j.asw.2010.05.002

Gebril, A., & Plakans, L. (2009). Investigating source use, discourse features, and process in integrated writing tests.

Spaan Fellow Working Papers in Second or Foreign Language Assessment, 7, 47–84.

Gebril, A., & Plakans, L. (2013). Toward a transparent construct of reading-to-write tasks: The interface between discourse features and proficiency. *Language Assessment Quarterly*, *10*(1), 9–27. doi:10.1080/ 15434303.2011.642040

Grabe, W. (1991). Current developments in second-language reading research. *TESOL Quarterly*, *25*(3), 375–406. doi:10.2307/3586977

Guo, L., Crossley, S. A., & McNamara, D. S. (2013). Predicting human judgments of essay quality in both integrated and independent second language writing samples: A comparison study. *Assessing Writing*, *18*(3), 218–238. doi:10.1016/j.asw.2013.05.002

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate data analysis* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

Hong Kong Curriculum Development Committee. (1975). *Provisional syllabus for Chinese Language (Form I-V)*. Hong Kong, PRC: Government Printer.

Hong Kong Curriculum Development Council & Hong Kong Examinations and Assessment Authority (HKCDC & HKEAA). (2007). *Chinese Language curriculum and assessment guide (Secondary 4 –6)*. Hong Kong, PRC: HKCDC & HKEAA. Retrieved from http://www.edb.gov.hk/attachment/tc/curriculum-development/kla/chi-edu/chi_lang_ final_5805.pdf

Hong Kong Curriculum Development Council (HKCDC). (1990). *Syllabus for Chinese Language (Secondary 1–5)*. Hong Kong, PRC: Government Printer. Retrieved from http://www.edb.gov.hk/attachment/tc/curriculum-develop ment/kla/chi-edu/ChiSyl_1990.pdf

Hong Kong Curriculum Development Council (HKCDC). (2001a). *Chinese Language curriculum guide (Secondary1-7)*. Hong Kong, PRC: Education and Manpower Bureau.

Hong Kong Curriculum Development Council (HKCDC). (2001b). *Learning to learn - the way forward in curriculum*. Hong Kong, PRC: Government Printer. Retrieved from http://www.edb.gov.hk/tc/curriculum-development/cs- curriculum-doc-report/wf-in-cur/index.html

Hong Kong Examinations and Assessment Authority (HKEAA). (2005). *HKCEE Chinese Language level descriptors and exemplars for standards-referenced assessment*. Hong Kong, PRC: HKEAA.

Hong Kong Examinations and Assessment Authority (HKEAA). (2012). *Hong Kong diploma of secondary education examination revised assessment framework for category A subjects (Chinese Language)*. Hong Kong: HKEAA. Retrieved from http://www.hkeaa.edu.hk/DocLibrary/HKDSE/Subject_Information/chi_lang/2013hkdse-clang.pdf

Kaiser, H. (1970). A second generation little jiffy. *Psychometrika*, *35*(4), 401–415. doi:10.1007/BF02291817 Kaiser, H. (1974). An index of factorial simplicity. *Psychometrika*, *39*(1), 31–36. doi:10.1007/BF02291575

Kaufman, A. S., Ishikuma, T., & Kaufman, N. L. (1994). A Horn analysis of the factors measured by the WAIS-R, Kaufman Adolescent and Adult Intelligence Test (KAIT), and two new brief cognitive measures for normal adolescents and adults. *Assessment*, *1*(4), 353–366.

Kline, R. B. (2005). *Principles and practice of structural equation modeling*. New York, NY: Guilford Press.

Knoch, U., & Sitajalabhorn, W. (2013). A closer look at integrated writing tasks: Towards a more focussed definition for assessment purposes. *Assessing Writing*, *18*(4), 300–308. doi:10.1016/j.asw.2013.09.003

Lee, Y. W. (2006). Dependability of scores for a new ESL speaking assessment consisting of integrated and indepen- dent tasks. *Language Testing*, *23*(2), 131–166. doi:10.1191/0265532206lt325oa

Lewkowicz, J. (1994, December). *Writing from sources: Does source material help or hinder students' performance*. Paper presented at the Annual International Language in Education Conference, Hong Kong, PRC (ERIC Document Reproduction Service No. ED386050).

McCrae, R. R. (1989). Why I advocate the five-factor model: Joint factor analyses of the NEO-PI with other instruments. In D. M. Buss, & N. Cantor (Eds.), *Personality psychology: Recent trends and emerging directions* (pp. 237–245). New York, NY: Springer.

McGhee, R. (1993). Fluid and crystallized intelligence: Confirmatory factor analysis of the Differential Abilities Scale, Detroit Tests of Learning Aptitude-3, and Woodcock-Johnson Psycho-Educational Battery - Revised. *Journal of Psychoeducational Assessment Monograph Series: Woodcock Johnson Psychoeducational Battery - Revised*, 20–38. Brandon, VT: Clinical Psychology Publishing.

Oxford, R. L. (1990). *Language learning strategies: What every language teacher should know*. New York, NY: Newbury House /HarperCollins.

Peverly, S. T., Ramaswamy, V., Brown, C., Sumowski, J., Alidoost, M., & Garner, J. (2007). What predicts skill in lecture note taking? *Journal of Educational Psychology*, *99*(1), 167–180. doi:10.1037/0022-0663.99.1.167

Plakans, L. (2008). Comparing composing processes in writing-only and reading-to-write test tasks. *Assessing Writing*, *13*(2), 111–129. doi:10.1016/j.asw.2008.07.001

Plakans, L. (2009a). The role of reading strategies in integrated L2 writing tasks. *Journal of English for Academic Purposes*, *8*(4), 252–266. doi:10.1016/j.jeap.2009.05.001

Plakans, L. (2009b). Discourse synthesis in integrated second language writing assessment. *Language Testing*, *26*(4), 561–587. doi:10.1177/0265532209340192

Plakans, L., & Gebril, A. (2012). A close investigation into source use in integrated second language writing tasks.

Assessing Writing, *17*(1), 18–34. doi: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. doi:10.1016/j.jslw.2013.02.003

Purpura, J. (1999). *Strategy use and second language test performance: A structural equation modeling approach*.

Cambridge, UK: Cambridge University Press.

Purpura, J. E. (1997). An analysis of the relationships between test-takers' cognitive and metacognitive strategy use and second language test performance. *Language Learning*, *47*(2), 289–325. doi:10.1111/0023-8333.91997009

Purpura, J. E. (1998). Investigating the effects of strategy use and second language test performance with high- and low-ability test-takers: A structural equation modelling approach. *Language Testing*, *15*(3), 333–379.

Purpura, J. E. (2013). Cognition and language assessment. In A. J. Kunnan (Ed.), *The companion to language assessment* (pp. 1452–1476). Hoboken, NJ: Wiley/Blackwell.

Read, J. (1990). Providing relevant content in an EAP writing test. *English for Specific Purposes*, *9*(2), 109–121. doi:10.1016/0889-4906(90)90002-T

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. doi:10.1080/15434303.2011.633305 Sawaki, Y., Stricker, L. J., & Oranje, A. H. (2009). Factor structure of the TOEFL Internet-based test. *Language Testing*,

26(1), 005–030. doi:10.1177/0265532208097335

Spivey, N. N. (1997). *The constructivist metaphor: Reading, writing, and the making of meaning*. New York, NY: Academic Press.

Stark, K. D., & Laurent, J. (2001). Joint factor analysis of the children's depression inventory and the revised children's manifest anxiety scale. *Journal of Clinical Child & Adolescent Psychology*, *30*(4), 552–567. doi:10.1207/ S15374424JCCP3004_11

Tabachnick, B., & Fidell, L. (2013). *Using multivariate statistics* (6th ed.). Harlow, UK: Pearson Education.

Taylor, S. (1996). Nature and measurement of anxiety sensitivity: Reply to Lilienfeld, Turner, and Jacob (1996). *Journal of Anxiety Disorders*, *10*(5), 425–451. doi:10.1016/0887-6185(96)00021-7

Tse, S. K. (2009). Chinese Language education in Hong Kong: Twenty five years of educational research in Hong Kong. *Educational Research Journal*, *24*(2), 231–255.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York, NY: Academic Press.

van Gelderen, A., Schoonen, R., De Glopper, K., Hulstijn, J., Simis, A., Snellings, P., & Ste-venson, M. (2004). Linguistic knowledge, processing speed, and metacognitive knowl-edge in first- and second-language reading comprehension: A componential analysis. *Journal of Educational Psychology*, *96*(1), 19–30. doi:10.1037/0022- 0663.96.1.19

Vandergrift, L. (1999). Facilitating second language listening comprehension: Acquiring successful strategies. *ELT Journal*, *53*(3), 168–176. doi:10.1093/elt/53.3.168

Watanabe, Y. (2001). *Read-to-write tasks for the assessment of second language academic writing skills: Investigating text features and rater reactions* (Unpublished doctoral dissertation). University of Hawaii, Manoa, HI.

Weigle, S. C. (2002). *Assessing writing*. Cambridge, UK: Cambridge University Press.

Weigle, S. C. (2004). Integrating reading and writing in a competency test for non-native speakers of English. *Assessing Writing*, *9*(1), 27–55. doi:10.1016/j.asw.2004.01.002

Weir, C. (1993). *Understanding and developing language tests*. Hemel Hempstead, UK: Prentice-Hall.

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. doi:10.1080/15434303.2012.750661

Yang, H. (2009). *Exploring the complexity of second language writers' strategy use and performance on an integrated writing test through structural equation modeling and qualitative approaches* (Doctoral dissertation). The University of Texas, Austin, TX. Retrieved from http://repositories.lib.utexas.edu/handle/2152/6619

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. doi:10.1002/tesq.6

Yu, G. (2013a). The use of summarization tasks: Some lexical and conceptual analyses. *Language Assessment Quarterly*, *10*(1), 96–109. doi:10.1080/15434303.2012.750659

Yu, G. (2013b). From integrative to integrated language assessment: Are we there yet? *Language Assessment Quarterly*, *10*(1), 110–114. doi:10.1080/15434303.2013.766744

Zhu, X. (2005). *A study of setting standards of Chinese Language assessment in reading, writing and integrated skills*. Technical Report (unpublished). Project No: CD/C/13-0309. Commissioned by Education and Manpower Bureau, HKSAR. Singapore: National Institute of Education, Nanyang Technological University.

Zhu, X. (2012). Performance of Putonghua listening skills of secondary school students in Hong Kong. In S. D. Chan (Ed.), *A study on Putonghua proficiency test for secondary school students* (pp. 59–108). Hong Kong, PRC: Chung Hwa Book Co.

Zhu, X. (2013). *Research and development project on developing students' language competency in the integrative use of language skills and strategies in Chinese Language education at the secondary level*. Technical Report (unpublished).

Project No: CD/CLE/73-1106. Commissioned by Education Bureau, HKSAR. Hong Kong: The Hong Kong Polytechnic University.

Zhu, X. (2014). *Assessment for learning: Principles and strategies*. Beijing, PRC: People's Education Press.

Zhu, X. (2015). Implementing integrated testing of Chinese Language in Hong Kong secondary schools: The current situation and improvement strategies. *Educational Research*, *36*(5), 114–121.

Zhu, X., & Wu, Y. L. (2013). Secondary school students' difficulties and learning expectation in integrated Chinese Language tasks in Hong Kong. *Education Journal*, *41*(1–2), 27–45.

Zhu, X., & Wu, Y. L. (2014). Strategies for secondary students to overcome difficulties of completing integrated writing tasks in Chinese Language. In A. D. Zhang (Ed.), *Effective teaching and joyful learning (III)* (pp. 21–34). Singapore, Singapore: Pearson Education.