
A comparative study of composing processes in reading- and graph-based writing tasks

HUI-CHUN YANG

Department of English Instruction, National Hsinchu University of Education, Hsinchu, Taiwan

Bio Data:

Hui-Chun Yang has a Ph.D. from The University of Texas at Austin and has been an instructor of English in the United States and Taiwan. She also had excellent opportunities to teach English to students at the primary, secondary, and post-secondary levels. She is currently an Assistant Professor at National Hsinchu University of Education in Taiwan. Her specializations include language assessment, second language reading and writing, and learner training.

Abstract

The study compares EFL writers' processes in composing reading-based writing (RW) and graph-based writing (GW) tasks developed for a university English proficiency exam. Think-aloud protocols and interviews of ten university-level nonnative English-speaking writers were collected to explore writers' composing processes. The results revealed that both types of the tasks require global comprehension of source texts as well as integrative manipulation of available information for writing. Some differences, however, existed across tasks and writers of varying score levels, with the RW tasks eliciting a more interactive and facilitative process than the GW tasks for the higher scoring writers. These results suggested that these tasks might measure different aspects of academic writing ability. Several considerations of the task constructs should apply in properly determining their use in a language test. The findings could be used to provide insights into the nature of RW and GW tasks and contribute to the validity of source-based writing tasks.

Keywords: writing assessments, integrated-tasks, reading-based writing, graph-based writing

Introduction

The ability to integrate sources into writing has been considered important for academic success (Campbell, 1990; Leki & Carson, 1997). Therefore, a plethora of university assignments have involved writing from multiple sources (Horowitz, 1986; Kirkland & Saunders, 1991). In the same vein, writing tasks requiring writers to compose from language input (e.g., reading passages, lectures) or visual input (e.g., graphs, charts, diagrams) have also been increasingly incorporated into the

assessment batteries of a number of language tests (e.g., Test of English as a Foreign Language - TOEFL, Canadian Academic English Language Assessment - CAEL, International English Language Testing System - IELTS, General English Proficiency Test - GEPT) as a means to increase test fairness (Feak & Dobson, 1996; Read, 1990) and foster positive washback effects on learning and teaching (Cumming, Grant, Mulcahy-Ernt, & Powers, 2004; Fox, 2004). Despite their widespread acceptance, criticisms have been leveled against the use of such tasks due to the fact that they may introduce a source of construct irrelevant variance into the assessment (e.g., Charge & Taylor, 1997; Xi, 2005). In this case, writers' reading comprehension and graph comprehension ability may play roles in affecting writing performance, making score use and interpretation difficult.

The present study explores a proposed inclusion of a RW task or GW task other than a writing-only task in an English proficiency exam administered to liberal arts and science majors enrolling in the course of Freshman English. The exam aims to determine students' preliminary English abilities and if further English for Academic Purposes (EAP) support is needed. This change is to respond to a rising call for more authentic writing tests that simulate real-life writing situations. In considering whether RW tasks or GW tasks should be included in the exam, it is necessary to explore the underlying constructs of these tasks. This study sought to investigate the similarities and differences of writers' processes elicited by the RW and GW tasks.

Review of relevant literature

Process in Language Testing

Since the late 1980s, language testers have started to consider test-takers' processes to clarify relationships between test tasks and the target language use context. According to Bachman (1990, 2002), investigations of processes involved may provide insight into the construct validity of a language test. Such information helps determine whether test-takers go through the processes expected by test designers or, in other words, if the test actually measures what it proposes to measure. Most process and strategy research has concentrated on test-takers' mental operations in response to selected-response items (e.g., multiple choice, drag-and-drop, cloze) in reading and listening comprehension tests (Anderson, Bachman, Perkins, & Cohen, 1991; Cohen & Upton, 2007; Douglas & Hegelheimer, 2006). Relatively little is known about how test-takers approach constructed-response items (e.g., writing) (Cohen, 1994). To gain a better understanding about the validity of source-based writing tasks, it is necessary to examine how writers interact with the source texts and how their processes may vary across tasks and writers.

Process on Source-based Writing

A number of previous studies have examined the processes proficient and less proficient writers use in L2 reading-based writing tasks by drawing on constructivist models of discourse synthesis (Spivey, 1984, 1990, 1997; Spivey & King, 1989) in which *organizing*, *selecting*, and *connecting* processes are identified. In *organizing*, readers/writers refer to the text organization to identify the overall ideas of the text,

and create a text of their own by rearranging chunks of information in the source text. *Selecting* is when readers/writers determine the most important chunks of information from a pool of content units, and subsequently incorporate the selected ones for their writing. During *connecting*, writers make connection between their background knowledge and content in the sources. Plakans (2009), for example, compared L2 writers' processes in two reading-based writing test tasks based on think-aloud protocols. The results showed that proficient writers used significantly more discourse synthesis processes, *connecting* and *organizing* in particular, than their less proficient counterparts. Using a similar approach, Asención (2004) compared processes used by native English speakers, advanced ESL learners, and EFL learners in completing summary tasks. She found that *organizing*, *selecting* and *connecting* occurred less frequently compared to monitoring and planning across groups. In addition to these operations of textual transformation, research on L2 reading-based writing has indicated three other key composing processes including goal-setting, revising, and monitoring (Esmaeili, 2002; Stein, 1990). Overall, these studies contribute an understanding of major reading and writing processes involved in reading-based writing and the connections between writers' L2 proficiency and their corresponding linguistic or cognitive processes.

Another type of source-based writing, GW tasks that include visual sources in the prompt (Bridges, 2010), has rarely been addressed in language testing literature. Fortunately, cognitive psychology theories offer frameworks of graph comprehension and interpretation that may shed light on the interactions between graphical inputs and test-takers. Previous research has identified three central processes that occur successively during graph comprehension: encoding a visually identifiable feature of a graph (e.g., a line sloping upward), interpreting that feature in relation to their knowledge about graphs (e.g., a rising line implies a mounting relationship), and associating specific graph referents to the graph feature (e.g., "human population is increasing") (Bertin, 1983; Carpenter & Shah, 1998; Kosslyn, 1989; Lohse, 1993; Pinker, 1990). Some studies analyzed the perceptual processes people use to make interpretations about a specific graphical format. Carswell, Emery, Lonon (1993) examined participants' processes in responding to a series of line graphs and found that participants constantly engaged in *global productivity* (i.e., an overall trend in a graph) and *local productivity* (i.e., an *x*- or *y*-axis reference, an interpretation for a specific part of a graph). In experiments where participants compared two wedges without values in a pie chart, Gillan and Lewis (1994) found that many participants mentally lifted one wedge over the other to make appropriate comparisons.

Despite the abundance of literature on graphical information processing, only a few studies have addressed graph interpretation in a testing context. Moreover, most of these studies have been conducted in the contexts of listening (Ginther, 2002; Gruba, 1997) and speaking assessments (Katz, Xi, Kim, & Cheng, 2004; Xi, 2005, 2010). Very little research, apart from internal IELTS validation reports (Bridges, 2010; Mickan, Slater, & Gibson, 2000), has addressed these processes involved in writing assessments. Using think-aloud protocols, Mickan, Slater, and Gibson (2000) investigated nine IELTS candidates' test-taking processes and identified three key processes that occurred successively during graph-based writing: *planning prior to*

writing, formulating text, and editing. Questionnaire results reported in Bridges (2010) revealed six processes commonly used to complete IELTS Academic Writing Task 1: macro-planning, organizing, micro-planning, translating, monitoring, and revising. The findings also showed that macro-planning (e.g., goal-setting, task examination) and monitoring occurred more frequently for skilled writers than less skilled writers. The questionnaire used in the study focused mainly on metacognitive self-regulation operations while interactions between graph reading and writing were rarely discussed.

The studies reviewed above show that composing patterns or process types, in general terms, are similar in reading-based writing and graph-based writing. Moreover, L2 proficiency is found to be an important variable that affects process. However, comparisons between processes involved in completing these tasks have not been investigated in previous work. Based on the need of the university English proficiency exam development and the review of prior research, the following research question was proposed: What similarities and differences exist between L2 writers' processes in composing reading-based and graph-based writing tasks?

Method

Participants

The participants were recruited on voluntary basis and were introduced to the study by reading a brochure describing the purpose and procedure of the study. The participants were ten full-time undergraduate students enrolled in the course of Freshman English. They have been studying English for at least ten years and had a wide range of English proficiencies based on their writing scores on the RW and GW tasks. Table 1 presents the participant profile.

Table 1
Participant Profile

Writer*	Gender	Age	Discipline	RW score	GW score
Yifen	F	19	Taiwanese	4	4
Jen	F	21	Public Health	3	4
Sam	F	19	Chemistry	4	4
Dayi	M	19	Japanese	2	3
Lee	M	19	Chemistry	3	3.5
Feng	M	20	Physical Therapy	2	3
Genna	F	20	Health Care Management	2	3
Jing	F	21	Public Health	2.5	4
Peiling	F	20	Health Care Management	2	2
Wei	F	19	Japanese	2	3

*All names are pseudonyms.

Tasks

The tasks that stimulate the academic writing skills of summarizing and synthesizing source materials were designed for the study and for the potential use in the English proficiency exam. Four argumentative writing tasks, one RW and one GW task for two environmental topics (i.e., global warming and ecotourism), were developed. The argumentative genre was selected because it was fairly common in most academic settings (Cumming et al., 2005). Following the suggestion of using more than one source text for source-based writing tasks (Lewkowicz, 1994), two short passages that present opposing viewpoints were included in each RW task. These passages were modified to be similar in text length, organization, and readability based on several criteria: specific main ideas and supporting details for an argument; Flesch Kincaid Grade level between 11 to 12; Flesch Reading Ease score between 40 to 60; and word count between 210 to 250.

Similarly, the GW tasks were created parallel to each other based on three rough standards: obvious trend changes, number of data points, and chart organization. Line graphs were found best in displaying x - y trends (Carswell, Emery, & Lonon, 1993; Shan, Mayer, & Hegarty, 1999) and supporting global-integration processing (Carswell, 1990; Hollands & Spence, 1992). On the other hand, pie graphs are good for depicting relative proportions of the data (Simkin & Hastie, 1987; Wilkinson, 1999). Line and pie graphs were thus selected in the development of source graphs in attempts to elicit global comparisons of two graphs rather than mere descriptions of x - and y -axis and data point values.

These tasks were reviewed by three EFL writing specialists and piloted on five potential participants who were also undergraduate freshman students in the university. Several issues including difficult vocabularies, idioms and slangs, sentences and graph structures, and task instructions were addressed before their actual use in the study.

Data Collection Procedures

Concurrent think-aloud verbal protocols, pre- and post-interview responses, and written products were collected to address the research questions. Think-aloud approach was used to gain access to the mental log of individuals performing an assigned task (Cohen, 2000; Ericsson & Simon, 1993; Green, 1998; Mickan, Slater, & Gibson, 2000). According to Krapels (1990), the technique provides valuable insights into L2 writers' cognitive operations during writing. Green (1998) also suggests that verbal protocols are a more direct means of "gathering evidence that supports judgments regarding validity than some of the other quantitative methods" (p. 3). By asking the writers to verbalize their thoughts when responding to the tasks, it was possible to investigate the cognitive processes by which writers transform ideas from sources into their own writing. Field notes detailing instances of writers' behaviors (e.g., underlining) and non-verbal expressions (e.g., frowning) were created to facilitate the interpretation of verbal reports. In addition, individual pre- and post-interviews were conducted to provide data for triangulation.

All data were collected in three sessions. The first session began with a brief orientation to the source-based tasks with sample tasks and an interview about writers' backgrounds, reading, graph reading, and writing experience. Then each

writer received instructions on thinking aloud. They watched a video demonstration of the think-aloud approach, and practiced responding to one sample RW and GW task different from the actual tasks. During the practice, writers were reminded to 1) verbalize their immediate thoughts instead of interpreting them, and 2) talk continuously if they fell silent for more than 20 seconds (Ericsson & Simon, 1993; Green, 1998; Olson, Duffy, & Mack, 1985). Also, given that some participants might be concerned about which language to use in the think-aloud session (Manchón, Murphy, & Roca de Larios, 2005), they were instructed to use any language, at any time, with which they felt comfortable. At the same time, feedback was provided to the writers until they became familiar with the technique (Cohen, 2000).

As soon as the writers were ready, they proceeded to the second session. During this session, each participant completed one RW task and one GW task. The study design was counterbalanced on topic order and task order to reduce any possible order effects (see Table 2). Although there was no time limit for completing the tasks, all participants managed to finish them within one hour.

The last interview session took place within two days after the writing session. This semistructured interview explored writers' approaches to complete the tasks and personal reactions toward the tasks. To thoroughly identify the processes involved and consider all sources contributing to the processes, unusual or unexpected comments and pauses were brought up for discussion.

Table 2
The Study Design

Task types	Topic: ecotourism	Topic: global warming
RW	1 st : Yifen, Jen	2 nd : Sam, Dayi, Lee
GW	1 st : Sam, Dayi, Lee	2 nd : Yifen, Jen
RW	2 nd : Jing, Peiling, Wei	1 st : Feng, Genna
GW	2 nd : Feng, Genna	1 st : Jing, Peiling, Wei

Data Transcription and Analysis

The analyses proceeded in four stages. The first stage involved analyzing the ten writers' composing processes and their interview responses. The audio-recorded verbal protocols were transcribed verbatim, and the transcripts were segmented into "idea units," described by Kroll (1977) as "a chunk of information which is viewed by the speaker/writer cohesively as it is given a surface form ... related ... to psychological reality for the encoder" (p. 85), for further analysis. The interview data were transcribed and analyzed thematically across interviews. Themes and patterns were identified and categorized to cross-reference with the verbal protocols.

Then two EFL writing specialists (including the researcher) with two to five years of teaching experience explored patterns and identified categories of processes in the data using line-by-line coding approach (Glaser, 1978). In the course of establishing the coding system (Table 3), two objectives were considered: 1) the coding system had to account for twenty sets of protocols as well as to allow comparisons between processes involved in completing the RW and GW tasks, and 2) process theories reviewed earlier were considered in creating categories. Once the categories had been defined, two coders coded a total of 20 protocols independently. In addition, to allow for identification of the composing sequence, the protocols were reexamined for the groupings of the processes and marked according to the composing phases.

Table 3
Coding System

Process	Description
Goal-setting	This involves checking, understanding, and interpreting task prompts and instructions.
Global planning	This involves identifying major ideas or trends in source materials based on writers' background knowledge about article and graph structures.
Local planning	For RW tasks, this relates to understanding source passages by breaking lexical items, phrases, or sentences into parts. For GW tasks, this involves reading data values or describing <i>x</i> - and <i>y</i> -axes.
Selecting	This refers to episodes in which L2 writers go through source materials and selectively draw relevant information from sources to support their writing.
Connecting	This is a process in which L2 writers search for a relationship in source materials.
Translating	This involves the transformation of thoughts and ideas into writing.
Revising	This is a process in which L2 writers adjust their written texts at word, sentence, or essay level as a result of monitoring.
Monitoring	This involves reflecting and checking on overall task progress and fulfillment and identifying mechanical issues such as spelling, punctuation, word choice, and syntax.

While the study was primarily qualitative in nature, quantifying the process data helps provide more information on the trends of the processes within and across tasks and writers. In the third stage, descriptive analyses including frequency counts, percentages, and central tendency statistics were performed to examine the possible differences and similarities which exist in L2 writers' processes when composing the RW and GW tasks.

In the fourth stage, the score pattern of each writer's essay was analyzed to examine the relationship between score and process. The same two experts thus rated the written texts ($n = 20$) separately. Prior to the actual scoring, a training session was held to provide scoring guidelines including the RW and GW scoring rubrics (see Appendix A) and anchor essays for each score level. Raters first reviewed the scoring rubrics for RW and GW tasks and clarified the traits for rating. Then they practiced rating five essay responses for the RW and GW tasks

respectively and discussed the scoring criteria and discrepancies between each level. As soon as raters reached an agreement on anchor essays of different score levels, they continued to score the entire set of essays. The texts were rated holistically on a scale of 0 to 5. Content, organization, language use were considered in rating the essays. The inter-rater reliability for the raters was found to be substantial for the RW tasks ($Kappa = .83, p < .001$) and GW tasks ($Kappa = .83, p < .003$). Because no two scores assigned by the raters differed by more than one score point, writers' overall writing performance was represented by an average of the two assigned scores.

Results and Discussion

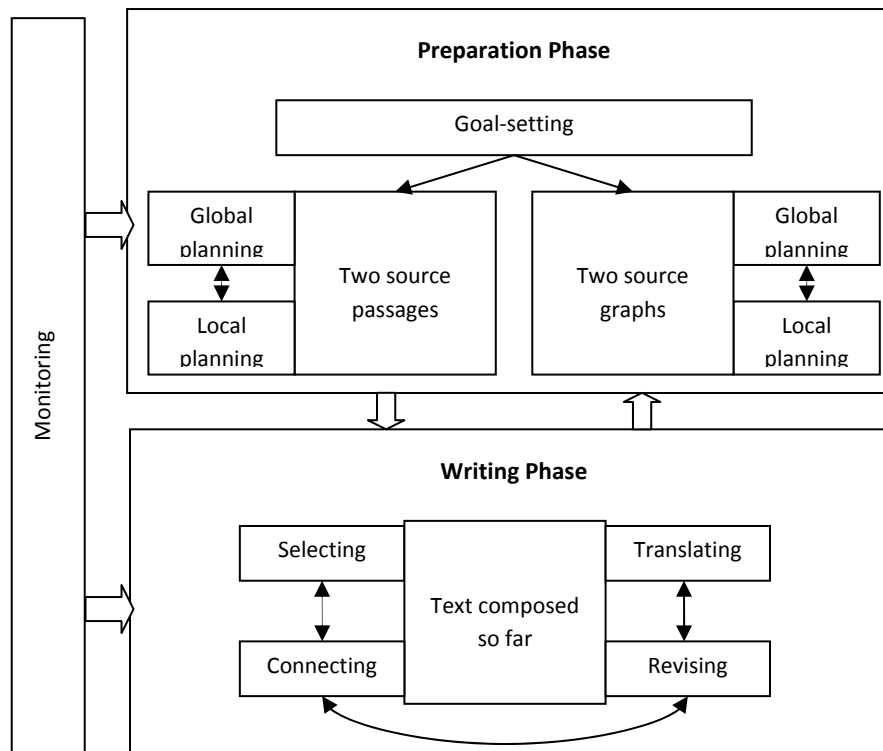
An L2 Source-based Writing Model

Based on the think-aloud protocols and interview data, a working model (Figure 1) of the writers' composing processes was developed. As Weigle (2002) suggested, L2 writing models should address the contexts in which writer-text interactions take place. This model is a context-specific schematic representation of L2 writers' processes involved in the RW and GW tasks. Although some individual differences were observed, two phases generally occurred for all writers: a preparation phase and a writing phase.

In the preparation phase, all writers read prompts, instructions, and then the source materials. Most of them read prompts and instructions at least once, and then reread some key words and phrases to understand the task goals and purposes further. After examining the task directions, they proceeded to read source passages and graphs using both *global planning* and *local planning*. In the writing phase, the writers reread a portion of source passages or graphs to select information for writing and for comparison between two source materials. When the pieces of information for writing were gathered, they composed their writing and evaluated what had been written. During the preparation phase, the writers read the source texts for the purpose of comprehension, while during the writing phase they reread the source texts for the purpose of summarization. Throughout the two phases, monitoring plays an important role in regulating the use of all processes.

In examining individual writer's process, some writers were found to engage in a more dynamic and constructive process than others. These writers (Sam, Lee, Yifen and Jen) appeared to be those who reported being interested in writing and having daily or weekly blog writing habits. For example, Sam said "It's just like what I often do for blogging. I wrote about what I read and if I am not sure about something, I went back and check for details ... I don't write well [sign] but I try very hard." Furthermore, Dayi and Genna who approached the tasks in a more linear fashion described having little experience with writing in their interviews.

Figure 1. Composing Processes for RW and GW Tasks



Composing Processes in the Source-based Writing Tasks

Table 4 presents the average percentage of each process during the preparation and writing phase across the RW and GW tasks. The subtotal segments coded as process for the two phases revealed that the writers devoted more time and efforts to the preparation phase than the writing phase during the RW tasks. This appeared to be the opposite for the writers responding to the GW tasks. It may suggest that textual sources, rather than graphic sources, place higher demands on writers’ cognitive resources for comprehension. The GW tasks, on the other hand, require more cognitive operations during writing than preparation. The following section compares the composing processes involved in the RW and GW tasks.

Table 4
Average percentage of each process for each type of task

	RW			GW		
	N	Mean (sd)	Range	N	Mean (sd)	Range
Preparation phase						
Goal-setting	60	.07 (.05)	12.7	59	.09 (.06)	17.2
Global planning	54	.05 (.04)	12.8	83	.12 (.05)	15.1
Local planning	354	.36 (.14)	50.5	91	.15 (.10)	31.1
Subtotal	468	.48	-	233	.36	-
Writing phase						
Selecting	174	.14 (.09)	25.7	103	.13 (.06)	22.5

Connecting	33	.03 (.03)	7.7	31	.03 (.03)	8
Translating	220	.22 (.07)	18.7	169	.25 (.09)	31.3
Revising	21	.02 (.02)	4.9	56	.05 (.06)	15.9
<i>Subtotal</i>	448	.41	-	359	.46	-
<i>Both phases</i>						
Monitoring	135	.11 (.09)	31.9	159	.17 (.11)	35.5

Note. N = the total number of think-aloud segment

a. *Preparation phase*: The preparation phase revealed some major differences between the RW and GW tasks. Such differences appeared in the average percentages of *local planning* and *global planning*. *Local planning* occurred at a high rate (36%) for the RW tasks, with participants trying to understand the meaning of source passages by focusing on word-level or phrase-level information. Some writers were found to continuously translating English words into Chinese to capture the main ideas of the passages. *Local planning*, however, occurred at a much lower rate (15%) during the GW tasks. On the contrary, *global planning* occurred at approximately double the rate during the GW tasks (12%) than the RW tasks (5%). As the interviews revealed, the readings provided writers with background ideas and organization for writing so little *global planning* was needed for the RW tasks. However, the GW tasks engaged writers in a greater level of rhetorical and structural planning for writing, suggesting that such tasks would be better in terms of making inferences about a writer's ability to apply logical structures to the content. The following are examples of processes *goal-setting*, *global processing*, and *local processing*:

1) '[reads the instructions] So summarize the ideas ... I am supposed to write a summary on two essays, okay, about one hundred fifty words, and here are two essays below. So summarize means ... to get main ideas. Main ideas from these two passages ... passage one and two. Ok.' (Yifen, RW)

2) '[reads the graphs] Zhe nian pai fang sheng gao (the emissions have gone up since this year) ... chi xu sheng gao dao liang qian nian (and keep going up to two thousand). Zhe bian zui gao (here is the highest) liang qian nian zhi hou kai shi xia jiang (and they went down after two thousand).' (Peiling, GW)

3) '[reads a sentence and translates some phrases into Chinese] From heat waves re lang (heat waves) to storms to floods shui zai (floods) to fires to massive glacial glacial ... melts, the global climate ... di qiu wen du ma (atmospheric temperature)?' (Dayi, RW)

b. *Writing phase*: During the writing phase, the top two processes that occurred most frequently were *translating* and *selecting*. *Translating* occurred at a highest rate in this phase for both RW (22%) and GW (25%) tasks. *Selecting* was the next most frequently observed process for the RW (14%) and GW (13%) tasks. The number of think-aloud segment revealed that writers were engaged in more *selecting* process when composing the RW tasks than the GW tasks. This may suggest a more

dynamic knowledge selecting and transforming process. Several writers described in their interviews that they carefully chose some pieces of information that represented the essence of the passages and deliberately neglected the others. The result indicates that the RW tasks can better elicit writers' ability to assess the values of information and make informed decisions on what to cite in their writing. Examples of processes *selecting, linking, translating, and revising* include:

4) '[rereads source graphs] Okay high point I need to write about high point. Let me look ... here ... a high point ... very high here in year ... two thousand the number of tourists is about ... about thirty thousand.' (Lee, GW)

5) '[rereads source passages] The first passage talks about ecotourism ... the second one also about ecotourism ... the same thing ... but they are different, different in some way. They have different points ... one agrees this is a good idea ... but the other don't think so.' (Jing, RW)

6) '[writes down a sentence] Many scientists' prediction is not very well. Prediction for the problem is not right.' (Dayi, RW)

7) 'The number of tourists are more than twenty thousand ... no shi tai (tense) cuo le (wrong) [crosses out a word] ... ying gai shi (should be) 'was' twenty thousand.' (Feng, GW)

c. *Monitoring in preparation and writing phase*: The monitoring process occurred at a higher rate during the GW tasks (17%) than the RW tasks (11%). For example:

8) '[rereads source graphs] I don't understand this graph [frowns]... why? [rereads instructions] Make connections? This graph and these two here. It says it's about CO2 emissions between these two years. How about here? One is nineteen-ninety and this one two thousand. Oh so here these two highs and these two years. Got it! I am so smart.' (Jing, GW)

Variation in Composing Processes by Performance Level

To explore variation in composing processes among writers, RW and GW essays were scored to compare processes used at different performance levels. The writers' scores ranged from 2 to 4 with all half-point scores rounded up. For the RW task, five writers were at level 2, three at level 3, and two at level 4. For the GW task, one writer was at level 2, four were at level 3, and five were at level 4. The descriptive statistics for ratings of RW and GW task performance (Table 5) show that writers' mean scores on the GW tasks, regardless of the topics, were generally higher than those on the RW tasks, suggesting that the RW tasks seem to be more challenging than the GW tasks.

Table 5
Descriptive Statistics for the Ratings of RW and GW Task Performance

Task types and topics	Mean scores	Minimum	Maximum	Standard deviation
RW				
Ecotourism	2.7	2	4	.84
Global warming	2.6	2	4	.89
GW				
Ecotourism	3.3	3	4	.44
Global warming	3.4	2	4	.89

When considering the range of each writer's composing processes, no consistent pattern is presented for essays scored as 4 and the other two levels; however, clear differences were found in the process means across score levels (Table 6). A comparison of the process means across three levels shows that writers at the highest level were engaged in approximately one third more and one fifth more processes than writers at the lower two levels for the RW and GW task respectively. Another difference lies in the score distribution. For the RW task, the scores cluster at a lower end whereas the opposite is true for the GW task. A further examination of individual writer's interview data reveal that three writers, Jen, Lee, and Jing who scored 3 in the RW tasks but 4 in the GW tasks had much experience with graphs. They stated that lots of their academic coursework involved analyzing graphs and converting data into graphs, which might indicate that writers' familiarity with graphs can play an important role in their performance on the GW tasks.

Table 6
Range and Means of Processes Used by Each Score Level (n = 10)

	RW Process					RW Mean	GW Process				GW Mean	
L 4	139	148				143.5	64	30	18	55	87	84.4
L 3	81	141	78			100	53	11	36	69		67.75
L 2	104	169	80	41	70	92.8	62					62

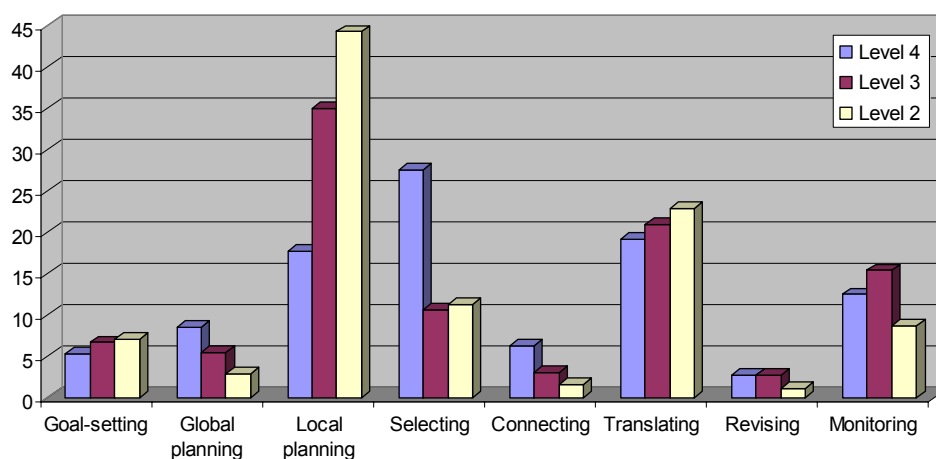
Note: L = score level. The numbers under the RW Processes and GW Processes columns refer to the total number of processes used by different score levels (e.g., writers who obtained four points for their RW tasks engaged in 139 and 148 composing processes during writing). RW mean and GW mean refer to the average of RW and GW processes used by each score level.

Previous research has shown that what distinguishes writers are the *range* as well as the *type* of processes involved in writing (Asención, 2004; Cohen, 1994). Figure 2 presents the occurrence percentage of each type of process engaged during the RW tasks. Several key patterns were found. The highest scoring writers engaged

in more global processes, such as identifying text structures and skimming for key ideas, which confirms previous reading-writing research (Cohen, 1994; Plakans, 2009). These writers also purposefully selected major ideas for writing and connected one piece of information with another from different source texts. The prevalent use of *global planning*, *selecting*, and *connecting* among higher scorers may suggest these were facilitative processes for the RW tasks.

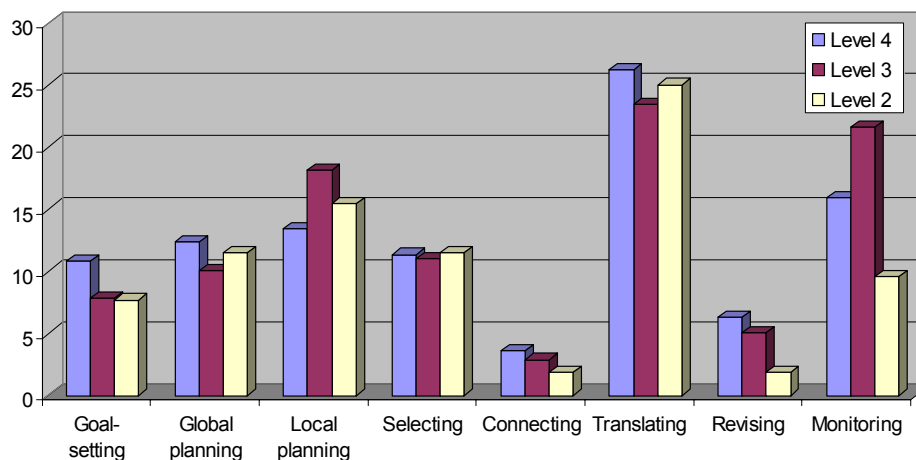
On the other hand, the writers at level 3 and 2 focused more on local-level information. They broke words and sentences into parts and resorted to translation frequently as a means to understand the whole texts, which might suggest that they found the source passages challenging than the higher scoring writers. Genna, Peiling, and Wei, who scored 2 in the RW tasks, had mentioned in their interviews of being 'uncertain,' 'confused,' and 'frustrated' about the source passages.

Figure 2. Percentages of Different Types of Composing Processes Used by Each Score Levels (RW)



Similar patterns were found for the GW tasks (Figure 3) while the differences between levels were fairly small compared to those in the RW tasks. Take *local planning* for example. Writers at level 4 in the RW tasks used this process much less frequently than the rest of the writers, but the difference was minor in the case of the GW tasks. This appeared to reflect the nature of these tasks. The RW tasks require a mastery of basic word- or sentence-level comprehension while the GW tasks do not. Yet what is not revealed in the figure is the language difficulties common to writers at level 2 and 3 in completing the GW tasks. Dayi, Genna, and Wei had indicated in their interviews that they felt the GW tasks were more challenging because they had to transform numerical data and visual trends into written texts. They often found difficulties in searching for appropriate trend-describing words (e.g., rise, fall, fluctuate) for writing. In contrast, Jen, Sam, Lee, and Feng, science majors who scored 4 in the GW tasks stated in their interviews that lots of their academic coursework involved analyzing graphs and converting data into graphs. These writers found their experience had helped them determine what to write and how to organize the content.

Figure 3. Percentages of Different Types of Composing Processes Used by Each Score Levels (GW)



Conclusion

The goal of the study was to investigate L2 writers' composing processes during source-based tasks to explore their underlying constructs. The two sets of tasks were designed and used in the study to measure academic writing skills. Concurrent verbal protocols were gathered to examine how writers created meaning through the transformation of source texts. The findings may enhance our understanding of the possible constructs of academic writing, and provide test designers and users more information for the development and interpretation of these test tasks.

The source-based writing model suggests that the writers approach both tasks with the types of problem-solving and discourse-synthesizing processes described in previous literature of source-based writing (Bridges, 2010; Hirvela, 2004; Spivey, 1997), which provides some evidence for construct validity (Messick, 1989) of these tasks. In addition, contrary to the purely linear (Rohman, 1965) and recursive view of writing (Flower & Hayes, 1981; Witte, 1985), the data reveal both linear and iterative nature of composing from sources depending on the writers' *procedural knowledge*, a repertoire to manage a range of thinking operations for the purpose of achieving the writing goals, as well as their *task environment*, the distance between writers' text composed so far and expectations they hold for themselves (van den Bergh & Rijlaarsdam, 2001).

The results on the proportions of processes engaged during the RW and GW tasks also reveal the nature of these tasks. The RW tasks minimally require a mastery of threshold reading comprehension skill at the sentence and paragraph levels. They are better in capturing writers' ability to evaluate sources and make selections for use in writing. Such ability is particularly important since authors almost always draw on ever-growing pools of information in academic writing. On the other hand, the GW tasks require writers to sequence the content in a unified logical structure as well as to compose using trend-describing vocabulary, which might pose some difficulties for writers who are not familiar with graphs.

When comparing the processes engaged by writers at different score levels, the data indicate that the highest scoring group showed a tendency to use facilitative

processes more frequently than the mid- and low-scoring groups during the RW tasks, a finding that can be linked to research on cognitive operations during reading-based writing (Connor & Kramer, 1995; Yang & Shi, 2003). Such differences, however, were not evident for the GW tasks. It seems that the RW tasks prompted more constructive processes than the GW tasks for the more proficient writers. These findings also reveal that these tasks may have been measuring different aspects of source-based writing ability. This assumption is further verified by the results that not all of the writers scoring high in the RW tasks scored high in the GW tasks. The three writers who scored high in the GW tasks but not in the RW tasks happened to be those who had much experience with graph interpretation and analysis. Clearly, graph familiarity and comprehension ability play roles in graph-based writing performance. As indicated in a number of studies on graph-based assessment tasks (Katz, Xi, Kim, & Cheng, 2004; Xi, 2005), caution must be taken due to the potential influence of graph familiarity on writers' processes and performance.

Given that GW tasks are designed to assess students regardless of their background knowledge on graphs, they were not considered to be included in the current English proficiency exam. However, GW tasks may have potential for use in assessing writing ability of science majors considering that visual literacy increasingly has been viewed as a prerequisite for understanding academic texts and graphicacy skills have started to be seen as part of the larger construct of academic writing (Hyland, 2006; Kress, 2003).

Several limitations of this study should be noted. First, the study only focused on four source-based writing tasks specifically developed for a university English proficiency exam. Task and topic effects may occur considering there are many variants of academic writing. In addition, although the think-aloud method can be effective for understanding writers' composing processes, they may affect or alter writers' thought processes and performances. Also, writers' processes would have been different if they participate in the actual proficiency exam where they would be limited in time. Finally, the study is exploratory in nature and only considers ten participants; thus, future studies incorporating more participants as well as participants with different backgrounds would provide greater insight into the nature of these tasks. In light of the restricted scope of this study, the results should be interpreted with caution.

References

- Anderson, N. J., Bachman, L. F., Perkins, K., & Cohen, A. D. (1991). An exploratory study into the construct validity of a reading comprehension test: Triangulation of data sources. *Language Testing*, 8(1), 41–66.
- Asención, Y. (2004). *Validation of reading-to-write assessment tasks performed by second language learners* (Unpublished doctoral dissertation). Northern Arizona University, Flagstaff, AZ.
- Bachman, L. F. (1990). *Fundamental consideration in language testing*. Oxford: Oxford University Press.
- Bachman, L. F. (2002). Some reflections on task-based language performance assessment. *Language Testing*, 19(4), 453–476.
- Bertin, J. (1983). *Semiology of graphics: Diagrams networks maps*. Madison, WI: University of Wisconsin Press.
- Bridges, G. (2010). Demonstrating cognitive validity of IELTS academic writing task 1. *Research Notes*, 42, 24–33.
- Campbell, C. (1990). Writing with others' words: Using background reading text in academic compositions. In B. Kroll (Ed.), *Second language writing* (pp. 211–230). Cambridge, UK: Cambridge University Press.
- Carpenter, P. A., & Shah, P. (1998). A model of the perceptual and conceptual processes in graph comprehension. *Journal of Experimental Psychology: Applied*, 4, 75–100.
- Carswell, C. M. (1990). *Graphical information processing: The effects of proximity compatibility*. Paper presented at the Human Factors Society 34th Annual Meeting, Santa Monica, CA.
- Carswell, C. M., Emery, C., & Lonon, A. M. (1993). Stimulus complexity and information integration in the spontaneous interpretation of line graphs. *Applied Cognitive Psychology*, 7(4), 341–357.
- Charge, N., & Taylor, L. B. (1997). Recent developments in IELTS. *ELT J*, 51(4), 374–380.
- Cohen, A. D. (1994). English for academic purposes in Brazil: The use of summary tasks. In C. Hill, & K. Parry (Eds.), *From testing to assessment: English as an international language* (pp. 174–204). London, England: Longman.
- Cohen, A. D. (2000). Exploring strategies in test taking: Fine-tuning verbal reports from respondents. In G. Ekbatani, & H. Pierson (Eds.), *Learner-directed assessment in ESL* (pp. 131–145). Mahwah, NJ: Erlbaum.
- Cohen, A. D., & Upton, T. A. (2007). 'I want to go back to the text': Response strategies on the reading subtest of the new TOEFL. *Language Testing*, 24(2), 209–250.
- Connor, U. M., & Kramer, M. G. (1995). Writing from sources: Case studies of graduate students in business management. In D. Belcher, G. Braine, & D. Dewhurst Belcher (Eds.), *Academic writing in a second language* (pp. 155–182). Norwood, NJ: Ablex Publishing Corporation.
- Cumming, A., Grant, L., Mulcahy-Ernt, P., & Powers, D. (2004). A teacher-verification study of speaking and writing prototype tasks for a new TOEFL. *Language Testing*, 21(2), 159–197.

- Cumming, A., Kantor, R., Baba, K., Erdosy, U., Eouanzoui, K., & James, M. (2005). Difference in written discourse in independent and integrated prototype tasks for next generation TOEFL. *Assessing Writing*, 10, 5–43.
- Douglas, D., & Hegelheimer, V. (2006, June 30). *Learner strategies and knowledge in TOEFL Listening Tasks: A validation study*. Paper presented at the Language Testing Research Colloquium, Melbourne, Australia.
- Ericsson, K. A., & Simon, H. (1993). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.
- Esmaeili, H. (2002). Integrated reading and writing tasks and ESL students' reading and writing performance in an English language test. *The Canadian Modern Language Review* 58(4), 599–622.
- Feak, C., & Dobson, B. (1996). Building on the impromptu: A source-based academic writing assessment. *College ESL*, 6(1), 73–84.
- Flower, L., & Hayes, J. R. (1981). A Cognitive Process Theory of Writing. *College Composition and Communication*, 32(4), 365–387.
- Fox, J. (2004). Test decisions over time: Tracking validity. *Language Testing*, 21(4), 437–465.
- Gillan, D. J., & Lewis, R. (1994). A componential model of human interaction with graphs: 1. Linear regression modeling. *Human Factors*, 36(3), 419–440.
- Ginther, A. (2002). Context and content visuals and performance on listening comprehension stimuli. *Language Testing*, 19(2), 133–167.
- Glaser, B. G. (1978). *Theoretical sensitivity: Advances in the methodology of grounded theory*. Mill Valley, CA: Sociology Press.
- Green, A. J. F. (1998). *Using verbal protocols in language testing research: A handbook*. Cambridge, UK: Cambridge University Press.
- Gruba, P. (1997). The role of video media in listening assessment. *System*, 25, 335–346.
- Hirvela, A. (2004). *Connecting reading and writing in second language writing instruction*. Ann Arbor, MI: The University of Michigan Press.
- Hollands, J. G., & Spence, I. (1992). Judgements of change and proportion in graphical perception. *Human Factors*, 34(3), 313–334.
- Horowitz, D. (1986). What professors actually want: Academic tasks for the ESL classroom. *TESOL Quarterly*, 20(3), 445–462.
- Hyland, K. (2006). *English for academic purposes: An advanced resource book*. New York, NY: Routledge.
- Katz, I. R., Xi, X., Kim, H., & Cheng, P. C.-H. (2004). *Elicited speech from graph items on the test of spoken English*. TOEFL research report no. 74. Princeton, NJ: Educational Testing Service.
- Kirkland, M. R., & Saunders, M. A. P. (1991). Maximizing student performance in summary writing: Managing cognitive load. *TESOL Quarterly*, 25(1), 105–121.
- Kosslyn, S. M. (1989). Understanding charts and graphs. *Applied Cognitive Psychology*, 3, 185–225.
- Krapels, A. R. (1990). An overview of second language writing process research. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 37–56). Cambridge, UK: Cambridge University Press.
- Kress, G. (2003). *Literacy in the new media age*. London: Routledge.

- Kroll, B. (1977). Combining ideas in written and spoken English: A look at subordination and coordination. In E. O. Keenan, & T. L. Bennett (Eds.), *Discourse Across Time and Space* (pp. 69–108). Los Angeles, CA: University of Southern California.
- Leki, I., & Carson, J. (1997). Completely different worlds: EAP and the writing experiences of ESL students in university courses. *TESOL Quarterly*, 31(1), 39–69.
- Lewkowicz, J. (1994). Writing from sources: Does source material help or hinder students' performance? In N. Bird (Ed.), *Language and Learning* (pp. 204–217). Hong Kong: Institute of Language in Education.
- Lohse, G. L. (1993). A cognitive model of understanding graphical perception. *Human-Computer Interaction*, 8, 353–388.
- Manchón, R. M., Murphy, L., & Roca de Larios, J. (2005). Using concurrent protocols to explore L2 writing processes: Methodological issues in the collection and analysis of data. In P. K. Matsuda, & T. Silva (Eds.), *Second language writing research: Perspectives on the process of knowledge construction* (pp. 191–205). Mahwah, NJ: Erlbaum.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–104). New York: American Council on Education & Macmillan.
- Mickan, P., Slater, S., & Gibson, C. (2000). Study of response validity of the IELTS writing subtest. In R. Tulloh (Ed.), *International English Language Testing System* (Vol. 3, pp. 29–48). Canberra, Australia: IELTS Australia.
- Olson, G. M., Duffy, S. A., & Mack, R. L. (1985). Question asking as a component of text comprehension. In A. C. Graesser, & J. B. Black (Eds.), *The psychology of questions* (pp. 219–226). Hillsdale, NJ: Erlbaum.
- Pinker, S. (1990). A theory of graph comprehension. In R. Freedle (Ed.), *Artificial intelligence and the future of testing* (pp. 73–126). Mahwah, NJ: Erlbaum.
- Plakans, L. (2009). Discourse synthesis in integrated second language writing assessment. *Language Testing*, 26(4), 561–587.
- Read, J. (1990). Providing relevant content in an EAP writing test. *English for Specific Purposes*, 9, 109–121.
- Rohman, D. G. (1965). Pre-writing: The stages of discovery in the writing process. *College Composition and Communication*, 16, 106–112.
- Shan, P., Mayer, R. E., & Hegarty, M. (1999). Graphs as aids to knowledge construction: Signaling techniques for guiding the process of graph comprehension. *Journal of Educational Psychology*, 91(4), 690–702.
- Simkin, D., & Hastie, R. (1987). An information-processing analysis of graph perception. *Journal of the American Statistical Association*, 82, 454–465.
- Spivey, N. (1984). *Discourse synthesis: Constructing texts in reading and writing*. Newark, DE: International Reading Association.
- Spivey, N. (1990). Transforming texts: Constructive processes in reading and writing. *Written Communication*, 7, 256–287.
- Spivey, N. (1997). *The constructivist metaphor: Reading, writing, and the making of meaning*. New York: Academic Press.
- Spivey, N. N., & King, J. R. (1989). Readers as writers composing from sources. *Reading Research Quarterly*, 24, 7–26.

- Stein, V. (1990). Exploring the cognition of reading-to-write. In L. Flower, V. Stein, J. Ackerman, M. J. Kantz, K. McCormick, & W. C. Peck (Eds.), *Reading-to-write: Exploring a cognitive and social process* (pp. 119–143). New York, NY: Oxford University Press.
- van den Bergh, H., & Rijlaarsdam, G. (2001). Changes in cognitive activities during the writing process and relationships with text quality. *Educational Psychology, 21*, 373–385.
- Weigle, S. C. (2002). *Assessing writing*. Cambridge: Cambridge University Press.
- Wilkinson, L. (1999). Graphs for research in counseling psychology. *The Counseling Psychologist, 27*(3), 384–407.
- Witte, S. P. (1985). Revising, composition theory and research design. In S. Witte (Ed.), *The acquisition of written language: Response and revision* (pp. 250–284). Norwood, NJ: Ablex.
- Xi, X. (2005). Do visual chunks and planning impact performance on the graph description task in the SPEAK exam? *Language Testing, 22*(4), 463–508.
- Xi, X. (2010). Aspects of performance on line graph description tasks: Influenced by graph familiarity and different task features. *Language Testing, 27*(1), 73–100.
- Yang, L., & Shi, L. (2003). Exploring six MBA students' summary writing by introspection. *Journal of English for Academic Purposes, 2*, 165–192.

Appendix A: Scoring rubrics for the RW and GW tasks

Score Level	Task Description
5	<p>A response at level 5</p> <ul style="list-style-type: none"> • is effective in selecting major information from two source texts/graphs to support one another and connecting relevant ideas • demonstrates unity, coherence, syntactic variety, and appropriate word choice • contains minor lexical or syntactical errors that do not interfere with meaning
4	<p>A response at level 4</p> <ul style="list-style-type: none"> • is effective in selecting and connecting major information from two source texts/graphs although some ideas may not be fully elaborated • demonstrates unity, coherence, syntactic variety, and appropriate word choice although it may contain few unclear connections or occasional redundancy • contains few lexical or syntactical errors that do not interfere with meaning
3	<p>A response at level 3</p> <ul style="list-style-type: none"> • contains some but not all major points from two source texts/graphs and the points are imprecisely or incorrectly presented or connected • demonstrates unity and coherence although it may contain few obscure connections and imprecise word choice • displays limited syntactic structures and vocabulary • contains some lexical or syntactical errors that occasionally obscure

	meaning
2	A response at level 2 <ul style="list-style-type: none">• contains limited relevant points from two source texts/graphs and they are significantly misrepresented• displays little organization or inadequate connections of ideas• contains inappropriate word choice• displays many lexical or syntactical errors that largely obscure meaning
1	A response at level 1 <ul style="list-style-type: none">• contains little or no relevant information from two source texts/graphs• is disorganized and underdeveloped• displays serious and frequent lexical and syntactical errors that make understanding of the writing unlikely
0	A response at level 0 <ul style="list-style-type: none">• contains copied words from the source passages• is written in a foreign language• is left blank