

Providing Source Materials in Second Language Writing Assessment Tasks: What Does It Do?¹

YUICHI WATANABE

Kanda University of International Studies

This article examines issues surrounding the use of source materials in writing assessment tasks. By reviewing relevant literature in first language and second language writing research, it first outlines potential benefits and problems of providing input materials in writing assessment tasks. Then, it discusses four aspects of student texts that are deemed important in evaluating source-based writing. The four aspects are content (origin of ideas and development of ideas), organization, function of integrated source information, and format of incorporated source materials.

1. Introduction

The importance of prompts in writing assessment, including impromptu essay tests, has long been recognized (Faigley, Cherry, Jolliffe, & Skinner, 1985; Hamp-Lyons, 1991; Hoetker & Brossell, 1986; Ruth & Murphy, 1988). A prompt includes both the topic and the task that a writer has to address. Kroll and Reid (1994) have classified the types of prompts into the following three categories (p. 233)

1. A “bare” prompt: In relatively direct and simple terms, it states the entire task for the candidate.
2. A “framed” prompt: A situation or set of circumstances is presented, and then a task is presented based on the interpretation of the frame.
3. A text-based or reading-based prompt: A passage of authentic (or adapted) reading material ranging in length from one paragraph to several pages is presented to the student, and the student is then asked to write an essay which demonstrates either his or her ability to

¹ This article is based on the author’s doctoral dissertation (Watanabe, 2001).

interpret the content of the reading or to use ideas in the reading by applying them in ways directed by the prompt.

Typically, the prompt in writing placement test does not include any input materials beyond a brief explanation of context (a bare prompt in the above). In recent years, however, serious limitations of this form of writing assessment have been pointed out (Feak & Dobson, 1996; Johns, 1991; Read, 1990).

2. Role of Source Materials in Second Language Writing Assessment

2.1 Rationale

Building a link to some reading and listening materials has often been adopted in large scale second language writing assessments (e.g., Test of English for Educational Purposes, English Language Testing Service, International English Language Testing Service, Ontario Test of ESL). Motivations behind the use of reading and listening input in writing assessment are mainly twofold. One is the recognition of the important connection between writing and other skills (i.e., reading and listening) in real-world academic tasks. Needs analysis studies conducted in Britain and North America have repeatedly pointed out that writing tasks assigned in the British and North American postsecondary institutions typically require the use of information from outside sources such as readings and lectures (Bridgeman & Carlson, 1983; Ginther & Grant, 1996; Hale et al., 1996; Weir, 1983; Wesche, 1987). Smaller scale investigations conducted in individual institutions also highlight the need to require the use of outside sources in ESL composition classrooms because that is what students need to do in their academic discipline courses (Leki & Carson, 1994; Leki & Carson, 1997). Exactly the same argument has been advocated for writing tasks used in assessment (Feak & Dobson, 1996; Hamp-Lyons & Kroll, 1997; Johns, 1991; Paltridge, 1992) based on the assumption that writing assessment tasks should be relevant to what students need to do in the real world.

The other major reason for providing source materials is to mitigate the effects of differences in writers' background knowledge about the assigned topic. One serious difficulty in prompt construction is the need to find a topic that will not bias against any particular groups of students due to their different degrees of familiarity with the topic and yet is interesting enough to elicit optimal performance. Providing readings on the assigned topic can have the effect of giving an equal chance to all the writers. Weir (1993) strongly holds this view: "By basing writing tasks on written and/or spoken text supplied to the candidates

or on non-verbal stimuli, it is possible to ensure that in terms of subject knowledge all start equally at least in terms of the information available to them" (p. 135). Read (1990) also advocates linking writing tasks to reading and listening tasks in assessment for the same reason and for the sake of simulating the process of academic writing tasks closely.

2.2 Types of input sources

In the above large scale tests of English for academic purposes, reading and listening input has been provided in separate subtests that intended to measure reading and listening skills. In OTESL, for example, a series of tasks designed to measure students' listening, reading, writing, and speaking skills are based on an audio-taped lecture and two reading passages all on a common theme (Wesche, 1987). Similar formats are used, as placement tests for example, in individual institutions as well (e.g., Paltridge, 1992). These tests, in which subtests are linked by a common theme but are designed to measure multiple-skills separately, are called integrated language tests (Lewkowicz, 1997).

Another type of writing test based on some reading and listening input is not intended to measure students' reading and listening ability separately. In these tests, the source materials are provided as part of a writing prompt and only one construct, writing ability (based on input materials), is measured. Some examples of this type of tests in published reports are Liu (1997), Read (1990), Lewkowicz (1994), and Feak and Dobson (1996).

Another way to classify writing tests based on input sources is by the number of channels (i.e., aural or written) and materials involved in the input. Some provide an audiotaped lecture and a reading (Liu, 1997; Paltridge, 1992), a lecture and two readings (OTESL in Wesche, 1987), one reading only (Task 2 in Read, 1990), two readings (Lewkowicz, 1994), or five readings (Feak & Dobson, 1996). The length of the material is another matter. Generally, due to the limited test administration time, if a greater number of input materials are involved, each material tends to be shorter than otherwise. This is particularly so if the input materials are provided solely for the purpose of writing assessment. But length also depends on the level of expected language proficiency of examinees. The 60-minute writing placement test for international undergraduate students in the University of Michigan (Feak & Dobson, 1996), for example, provides five short quotes on a theme, each quote ranging from 80 to 100 words in length. A writing placement test at the University of Illinois at Urbana-Champaign (Liu, 1997) has a time limit of 50 minutes (for reading and writing) and is based on a video

lecture of between seven and eleven minutes and a reading passage of from 600 to 1000 words.

This diversity in the input materials reflects different emphases in the aspects of academic writing tasks that test developers attempt to simulate in their tests. It also indicates different skills necessary to accomplish real-world tasks that the tests are designed to assess. Because of inherent logistic limitations, the number of types of academic writing tasks to simulate and skills to measure are quite limited in any single-time writing tests. Within a limited time (usually between one and two hours), the tests need to give students an optimal chance to demonstrate the skills that testers intend to measure. Those tests that have students listen to a lecture and read a passage first before they write an essay apparently attempt to simulate typical academic writing tasks, such as essay tests and report writing. These tasks require students to interpret information in both lectures and their textbooks accurately and critically, and select and organize relevant information to fulfill specific goals. On the other hand, writing tests that include multiple reading passages in their prompts, as in the writing placement test described by Feak and Dobson (1996), focus on synthesis of relevant information from diverse sources. In report writing assignments, students need to decide which, among many sources, are relevant for their report and which information to include; they need to discover connections between pieces of information across different sources, identify similarities and differences, and perhaps resolve apparent contradictions. Finally, they need to integrate relevant information and organize it to achieve their purpose in writing. Even in essay tests in content courses this type of prompts are commonly found (Feak & Dobson, 1996).

2.3 Task difficulty

Does the provision of source materials make a writing task easier or more difficult for second language writers? Very few studies have investigated this topic. A study by Lewkowicz (1994) has compared the scores of two groups of ESL students in Hong Kong ($N = 75$). One group was provided with two reading passages on the test topic and the other was not. No statistically significant difference was found between the mean scores of these two groups. Most students who first read two reading texts on a topic wrote more number of ideas than those who didn't; however, they did not elaborate on each idea as much as the writing-only students. In addition, students with readings plagiarized from the source texts extensively. In first language writing assessment research, Brown,

Hilgers, and Marsella (1991) and Smith and his colleagues (1985) have examined this question. Brown et al. (1991) conducted two studies using a repeated-measures design ($N = 1702$ and 1750). Their data came from several administrations of a freshman composition placement test in which students wrote two essays: a personal essay (without sources) and a response essay (with one or two reading passages). Neither of the studies found significant differences between mean scores from the two types of prompts. Smith et al. (1985) have compared three prompt structures: an open structure (no sources), a response structure (with one reading passage), and another response structure (with three passages). They tested each prompt structure with three writing proficiency groups: students in basic writing courses ($n = 61$), in general writing courses ($n = 73$), and in advanced writing courses ($n = 42$). Thus, they had a three by three between-groups design ($N = 176$). All the essays were rated on a four-point holistic scale and results are summarized in Table 1 below.

Table 1
Means and Standard Deviations of Holistic Scores by Prompt Structure and Proficiency Group (Adapted from Smith et al., 1985)

	Open Structure		With One Passage		With Three Passages	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Basic Writer	2.20	.52	2.27	.63	2.37	.60
General Writer	2.32	.69	2.36	.49	2.92	.56
Advanced Writer	3.43	.51	2.71	.83	2.93	.62

On the open structure prompt the advanced group was significantly better than the two lower proficiency groups. On the response structure with three passages, the basic writers were significantly lower than the upper two groups. On the response structure with one passage, the three proficiency groups performed similarly; only the difference between the basic writers and the advanced writers reached statistical significance. A similar pattern was found in the number of words produced under each structure, i.e., fluency. Smith et al. speculate some possible reasons that the prompt with only one passage failed to distinguish the three proficiency groups (p. 81):

- (1) The length of the passage (206 words) cued them to produce about the same length.
- (2) The single passage did not give them enough evidence, perhaps even less than they could create themselves. Thus, this [prompt structure] actually disables them.

(3) [This prompt structure] may resemble what students typically see on tests in content areas and on placement and diagnostic examinations, perhaps causing them to respond with what might be called a standard response (what some have termed ‘theme-writing’).

Recently, Norris, Brown, Hudson, and Yoshioka (1998) have proposed a framework for determining task difficulty in second language performance assessment. Based on Skehan's (1996) work, their framework focuses on three major task components that are believed to affect task performance in second language: code complexity, cognitive complexity, and communicative demand. The first component, *code complexity*, represents (1) linguistic difficulty of the information that needs to be understood and produced (called 'range'), and (2) the number of different input sources. The second component, *cognitive complexity*, represents the complexity of cognitive operations that the learner has to engage in to perform a task. This component has two variables: input/output organization and input availability. The first one deals with the degree of selection and re-organization of input and/or output that needs to be made based on some external criteria. The second variable indicates whether the information necessary to perform a task is readily available or must be searched and obtained. The third component, *communicative demand*, has two variables: mode, that is whether a task achievement requires some production or not, and response level, which represents the immediacy (e.g., on-line) of response that a task demands. According to this framework, providing input materials in writing could increase the difficulty of an assessment task. However, as Norris and his colleagues would agree, the presence of multiple sources of input in itself does not make a particular task more or less difficult. It depends on what is to be done with the source materials. If a writing task requires the learner to simply read provided materials for background information, and does not particularly require her to do anything with them (i.e., scoring criteria does not specify manipulations or incorporation of the source materials), the task may be easier than without such materials. On the other hand, if a successful task completion necessitates comparing, contrasting, and evaluating main points in the sources, that might be more difficult than doing the same with only a short-paragraph length description of an issue to be analyzed. Both the number of different input materials and what to do with them (including selection and re-organization) affect task difficulty. Obviously, code complexity represented by the linguistic difficulty of input materials will be another important factor in the difficulty of writing task with

input materials. As the studies by Lehner, Wolfe-Quintero, Tripodi-Segade, and Hilgers (in press) and Spack (1997) have shown, one of the biggest obstacles for ESL students in higher education is reading long and difficult materials.

2.4 Problems in source-based writing tasks

Although source-based tasks in writing assessment have an intuitive appeal in simulating real-life academic writing tasks, some serious concerns have also been raised. One of them is over-reliance on and misuse of source materials, most notably plagiarism (Bloch & Chi, 1995; Campbell, 1990; Currie, 1998; Leki & Carson, 1997; Lewkowicz, 1994; LoCastro & Masuko, 1997, March; Pennycook, 1996). Unattributed borrowing of words and ideas from source materials is not limited to second language writers. It is one of the most often discussed issues in the literature on teaching academic writing to L1 English (i.e., native speakers of English) “basic” writers as well. The source of this problem, particularly for second language writers, is complex. It is caused by the combination of unfamiliarity with the Western rhetorical expectations in academic writing, the deficiency in linguistic proficiency to manipulate source texts appropriately, and the lack of time (Bloch & Chi, 1995; Currie, 1998; Leki & Carson, 1997; Lewkowicz, 1994; LoCastro & Masuko, 1997, March).

This problem is often used as an argument against the use of reading materials in writing assessment. Read (1990) states: “It is this kind of problem [i.e., extensive copying from source texts] that is often used as an argument against using a linear text—as distinct from a diagram, or a graph—as stimulus material in a writing test” (p. 115). The IELTS, for example, has discontinued a link between its reading and writing subtests partly because their examinees relied too heavily on the reading materials when they wrote (Charge & Taylor, 1997). Up till then, IELTS test takers worked on a reading section first and were given a writing test based on the readings.

However, it is also possible to argue that reading-based writing tasks can be used to distinguish the student writers who can make appropriate and effective use of source texts and those who cannot. In defending his writing assessment task with a reading passage (Task 2), Read states:

Many of the students, especially those who are weak at writing, have a tendency to plagiarize from their sources, and this writing task gives an indication of whether they have the ability to express the ideas from the text in their own words. (p. 115)

The problem of over-reliance on source texts and plagiarism will be discussed again later in the review of text analysis literature.

Another potential problem involved in the use of reading texts in writing assessment tasks is confounding of skills. If students did not do well in a writing test based on some readings, was this due to their low writing ability or low reading ability or both? This is another concern that led IELTS to the removal of the link between their writing module and reading module in 1995 (Charge & Taylor, 1997). However, as was described earlier with respect to task difficulty, the difficulty of reading-based writing tasks depends on the difficulty of the reading materials as well as what examinees are asked to do with the materials. In integrated language tests like the old IELTS, reading passages need to be difficult enough to measure examinees' reading proficiency. On the other hand, when reading materials are used only as input in a writing task, it does not have that restriction. On the contrary, they can be too short and too unchallenging as materials in a reading test for the examinees. If this is the case, perhaps it will be possible to create reading-based writing tasks that are appropriate for the reading proficiency of intended students; then, confounding of skills will not be a significant concern when interpreting test results. Nevertheless, it does not mean that in reading-based writing assessment the reading proficiency does not play a role even when supplied reading materials are relatively easy for the students. Writing from sources requires higher order, as opposed to decoding linguistic codes, reading skills, such as critical reading, finding similarities and differences in different authors. In this sense the two skills are inseparably entwined in the performance of source-based writing tasks. Some first language reading and composition studies have compared more proficient and less proficient readers writing from sources (Kennedy, 1985; Spivey, 1984; Spivey & King, 1989). In these studies better readers produced better compositions. However, no similar studies exist in second language literature. Some studies have explored relationships between reading ability and writing ability in second language (e.g., Carson, Carrell, Silberstein, Kroll, & Kuehn, 1990) but none has examined the role of reading ability in composing from sources.

Another concern about the use of source-based writing tasks in assessment is that scoring them might take a longer time than scoring non-source-based texts. This is supposedly due to the need to refer to the source texts repeatedly to check how well or how poorly (including plagiarism) the student used the source texts (Read, 1990). However, no empirical studies have documented this so far. Neither does there seem to be any study that has compared the reliability of both

types of prompts. Several studies have used both types of prompts (Brown et al., 1991; Lewkowicz, 1994; Smith et al., 1985), but they all report one inter-rater reliability value for both tests combined. The only exception is a study by Read (1990). He examined the reliability and validity of one “guided” task and two “experience” tasks. In the guided task, students were given a table that described features of three grammar books. Their task was to explain which book was the most suitable for them. In the first experience task, students were given a 600-word passage about the process of steel-making and 25 minutes to take notes. After that the passage was collected and the students were asked to use the notes to give their own account of the process of steel-making in 30 minutes. The second experience task was more extended. Three key articles on a topic were used in combination with various activities in classes over five days. Then, students were given an examination-type question and wrote an answer in 40 minutes. All the essays were scored with a 6-point holistic scale, each scored by two teachers. The average correlation between two raters for each task was .75 for Task 1, .77 for Task 2, and .73 for Task 3.² Judging from these figures, it seems that the inter-rater reliability of rating source-based writing is comparable to that of rating non-source writing. However, Read cautions against using these figures to make a judgment about relative reliability of these three tasks. Because different teams of raters were involved in rating each task, it is not possible to make a direct comparison between the three reliabilities. There is a need for more controlled studies that allow such a comparison.

3. How Are Source Materials Used in Student Texts?

3.1 Text features

Numerous studies have attempted to determine what features in compositions affect the raters’ holistic judgment of their quality (e.g., Breland & Jones, 1984; Freedman, 1979; Grobe, 1981; Stewart & Grobe, 1979 in first language literature and Chiang, 1999; Sweedler-Brown, 1993; Vann, Lorenz, & Meyer, 1991 in second language literature). But the compositions investigated in these studies were not based on source materials. There have been hardly any similar attempts to investigate the relationships between the features of source-based writing and raters’ evaluation of the writing (exceptions are Campbell, 1987, 1990 and Lewkowicz, 1994). However, there is a body of knowledge as to how students,

² Using Spearman-Brown Prophecy Formula, the reliability of the summed ratings would be .86 for Task 1, .87 for Task 2, and .84 for Task 3.

mainly native speakers of English, use source materials to construct their compositions.

Serious attempts to analyze the source-based composition process and product began with the seminal work of Spivey (1983; 1984). Since then, a number of studies in first language reading and writing research have examined the process and the product of source-based composition (Ackerman, 1991; Durst, 1987; Flower et al., 1990; Greene, 1993; Hilgers & Stitt-Bergh, 1993; Kennedy, 1985; Mathison, 1996; Spivey & King, 1989). On the other hand, in second language literature, only a few studies have analyzed the student compositions based on reading (Campbell, 1987; 1990; Johns & Mayes, 1990; Lewkowicz, 1994).

The above studies on source-based composition have mainly focused on four aspects of student texts: content, organization, function of incorporated source information, and format of incorporated source information. Specifically, the main questions addressed in these studies were:

1. What is the proportion of the source information and the writer's own ideas in students' texts?
2. How are the source information and the writer's own ideas integrated and organized?
3. What is the function of the incorporated source information? (e.g., support for the writer's own argument)
4. What is the format of the incorporated source information? (e.g., direct quotes, paraphrase, summary with proper documentation/citation)

In the following review, these four aspects will be examined because they are considered to be important elements in the assessment of source-based writing (see for example, the rating rubric used in a source-based writing placement test in Liu, 1997). The questions deal with the issues of how source information is and should be selected, transformed, integrated with the writer's own ideas, and organized, following academic conventions (including documentations) to accomplish a task successfully.

3.2 Content

Origin of information. Unless the purpose of writing is the summary of source materials, the writers in source-based writing tasks are expected to integrate their ideas with the information in the source texts. In many writing assignments including research papers, the expected focus of writing is the writer's own idea, perspective, or analysis of a particular topic. The information from outside

sources is selected, transformed, and organized to serve the writer's purpose (Spivey, 1990). However, as described earlier, one of the often-mentioned problems with source-based writing tasks is students' over-reliance on source texts (e.g., Charge & Taylor, 1997; cf. Read, 1990). This problem is not limited to second language writers (e.g., Higgins, 1993; Kennedy, 1985).

Ackerman (1991) and Greene (1993) have used "content unit" as a unit of analysis to investigate the origin of information in their students' texts. Content units are clausal units similar to Kroll's (1977) idea units. The target of their investigations was the effect of the writer's prior knowledge (Ackerman) and task (Greene) on the proportion between the writer's original ideas and the information taken from sources. In Ackerman's study, forty English as a first language graduate students in psychology and business were randomly assigned to one of the two source-based writing topics: rehearsal in memory or supply-side economics. Half of the participants wrote a synthesis on the topic related to their major (high-knowledge); the other half wrote on the other topic (low-knowledge). The source materials in each task were four intact passages. The mean length of the passages was 1015 words for supply-side economics and 804 words for rehearsal. The source texts and students' texts were first parsed into content units. Then, the content units in the students' texts were classified into three types based on their origin and format: new (from the writer's prior knowledge), borrowed-explicit (from a source text and the exact words or a close paraphrase were used), and borrowed-implicit (from a source text). Ackerman found that the compositions in the high-knowledge condition contained more "new" content units (44% of the total) and fewer "borrowed-explicit" units (12% of the total) than in the low-knowledge conditions (22% and 27% respectively). This indicates that student writers' dependence on source texts is influenced by the amount of background knowledge about the topic. The number of "borrowed-implicit" remained similar across the two conditions (51% in the low knowledge condition and 44% in the high-knowledge condition), and was much higher than "borrowed-explicit." Furthermore, overall (i.e., when the results of the two conditions were combined) although the majority of information (66%) came from the source texts, a substantial proportion (34%) came from students' original ideas. Ackerman attributes these results to the expertise of the graduate students as readers/writers.

Using a stratified random sampling procedure, Greene (1993) assigned fifteen upper-division undergraduate students in a social history class to one of

two tasks, writing an informational report or a problem-based essay.³ In both tasks the students were given ten days to read six texts on the effects of the European Recovery Program (Marshall Plan) and to write a paper in three to five pages. As in Ackerman's study, the source texts and student texts were parsed into content units. Then, the content units in student texts were identified as either "borrowed" (taken from one of the source texts) or "added" (new information not found in any source texts). It was expected that problem-based essays would contain more overall content and a lower proportion of borrowed information than reports. However, the results indicated that although problem-based essays included more content units than reports, there was no significant difference in the proportion of borrowed information between the two types of writing. The mean proportion of borrowed content units was .62 ($SD = .16$) for problem-based essays and .71 ($SD = .19$) for reports.

Although Ackerman and Greene examined the proportion of ideas taken from sources, they did not investigate which sources were used most often. Lewkowicz (1994) has looked into both questions in her study of the effect of providing source texts on student essays. Two groups of first-year undergraduate students in Hong Kong wrote on the identity of Hong Kong in a 50-minute class. One of the groups had been given 25 minutes to read two passages on the topic and take notes one or two days before; the other groups had not. The passages were extracts taken from an article (745 words) and a book (805 words). The number of "points" was counted in the passages (22 in the first text and 11 in the second) and the reading-based group's texts. The 'points' were "linguistic propositions in that they are assertions that contain truth value" (p. 208). The ratio of points in student texts that originated in the source texts and those that were generated by students was approximately four to one. The proportion of the student-generated points (i.e., about 20%) is lower than the proportion of "added" content units (29%) in Greene's informational reports (Greene, 1993) and close to "new" content units (22%) in Ackerman's low-knowledge condition (Ackerman, 1991). However, since the unit of analysis was different among the three studies, particularly between Lewkowicz's study and Ackerman's and Greene's studies, this comparison needs to be interpreted with caution. As to the question of which source text is used more in student texts, Lewkowicz's found that the ratio between the points incorporated from the first text and the second text was about nine to two, much larger than the ratio of the total number of

³ Greene does not indicate whether any of the students were ESL students.

points in each text, which was two to one. The reason for this much heavier reliance on the first text seemed that the students had found it easier to lift materials directly from the first text than from the second one. Through interviews of a subgroup of students ($n = 11$), Lewkowitz found that to some students the first text provided more readily copyable materials whereas the second text did not as “the relevant ideas in the latter were scattered and the text included a number of points that were not related to the question” (p. 211). Direct lifting of materials from source texts will be discussed again later in relation to the function and the format of incorporated source information.

Development of ideas. If two groups of students write on the same topic and one group was given some reading passages on the topic, will their ideas be more developed than the other group on account of the relevant information they can utilize? This question was also examined by Lewkowitz (1994). She found that the group who read two reading passages wrote more points than the other group ($M = 10.6$ vs. 8.6 , $p < .05$) but the total length of their essays was not significantly different from the other group ($M = 435$ vs. 397 , n.s.). According to Lewkowitz, this indicates that each point introduced in the source-based essays was less developed than the non-source essays. However, the lack of statistically significant difference in the mean essay length seems to have been due to a large variation in both groups ($SD = 122$ with source texts vs. 101 without source texts). Therefore, the extent of the difference in idea development between the two groups, if there was any, is not clear.

3.3 Integration/Organization

The organization of content in multiple-source-based writing is closely related to how the writer conceptualizes the role of the source texts. The role of source texts depends on the task requirement and the writer’s task representation. In some cases, writers may read the source texts just to stimulate their own thoughts; once they have conceived an idea to write about, they may not have responsibility to respond to any points in the source texts. In other cases, the task may be perceived as one calling for a summary of all the key information in the source texts. Or the purpose of writing for some writers may be to find a common thread that runs across diverse sources and write a synthesis based on the thread. Or the purpose for some other writers may be to select and use whatever information is relevant to advance their own agenda in their text. These purposes have been called “organizing plans” by Flower and her colleagues (Flower et al., 1990). Organizing plans are important because they “structure

what [is] being read and to structure the writer's own text" (Flower, 1990b, p. 43). An intended organizing plan may not always be implemented successfully in the writer's text. For example, what was intended as a synthesis by students may be perceived as a summary by instructors. The organizing plan perceived by the instructors has an important implication because it can affect their judgment of the quality of the writing. Flower (Flower, 1990b) argues:

The logic of the organizing plan helps writers to create a (to them) coherent text structure and to signal that sense of coherence to the reader in places such as introductory paragraphs, transitions, topic sentences, and conclusions. Because the perception of coherence has such impact on the way instructors evaluate a paper, the organizing plan can influence students' grades. This is especially true if instructors value some organizing plans, such as synthesis or interpretation, as more intellectually significant than [sic] others, such as summary. (p. 44)

In an attempt to find out what organizing plan students select and realize in source-based writing tasks, Flower and her colleagues asked seventy-two college freshmen to read five short notes on time management for the express purpose of making a "comprehensive statement about this subject" (Flower, 1990a:26). Flower et al. identified seven organizing plans in student texts. They are summary, review and comment, isolated main point or conclusion, frame, free response to the topic, synthesis (controlling concepts), and interpretation for a purpose of one's own (see Figure 1 for the definition of each plan). The majority of the student texts exhibited the second, the third, and the fourth plans. In fact these three categories had been a single category called Review and Comment in their exploratory study conducted earlier. Flower et al. found that student texts sometimes sent conflicting signals about their structure that confused readers. For example, "the introductions might announce a plan to argue or discuss information, but the bodies of the papers would use coherence-building strategies that matched a quite different task interpretation, such as summarizing the information" (Kantz, 1990, p. 76). These coherence-building strategies gave false cues about the actual text structure. Based on this finding, Flower et al. created two subcategories within their original Review and Comment category: Isolated main point or conclusions, and Frame.

Category 0: Summary

States gist or selected ideas from the source text.

Category 1: Review and comment

Combines a summary or selective review of material from the source text with commentary or additions by the writer.

Category 2: Isolated main point or conclusion

A subcategory of review and comment containing a statement that is signaled as main point or a conclusion but does not appear to control the structure or selection of content in the body of the text.

Category 3: Frame

A subcategory of review and comment containing a statement that frames the review with a vague, highly general, or obvious introductory statement.

Category 4: Free response to the topic

Discusses the topic with little reference to information from the source texts.

Category 5: Synthesis (Controlling Concepts)

Organizes a discussion (which draws on source materials) around a unique (nonobvious) controlling concept.

Category 6: Interpretation for a purpose of one's own

Organizes a discussion (which draws on source materials) around a unique and apparent rhetorical purpose (beyond the purpose to summarize, comment, synthesize, etc.)

Figure 1. Organizing plan categories.

Note. From Kantz, 1990, pp. 77-78.

This taxonomy can be useful for identifying the structure of texts written in source-based writing tasks. As Flower et al. acknowledge, the seven categories do not necessarily form a hierarchy of text quality. Depending on a particular task, a summary may be expected over an interpretation. Yet in their task, Interpretation for a Purpose was considered most desirable, followed by Synthesis. Review and Comment varieties or Summary were not what instructors expected from the students. In writing assessment research, Hilgers and Stitt-Bergh (1993) used a modified version of this scheme to compare essays written in response to a source text by two groups of students at freshman writing placement tests. One of their results indicated that the texts of a more recent

group of high school graduates exhibited more of interpretation for a purpose of one's own than an earlier group. Hilgers et al. consider this as a sign of improved skills of incoming high school graduates. No similar studies with second language writers exist. A taxonomy similar to the one used by Kantz and Hilgers et al. may be used to investigate organizing plans of second language writers composing from multiple sources.

3.4 Function of incorporated source information

When writers incorporate information from sources, what function does it play in their text? Is a certain function more valued by instructors than others? For example, if some information in a source text is used as a support for the writer's assertion, will it be considered as a more effective use of a source than when it is used to introduce a new point? What function do most incorporations of source information play in student texts? In the study described earlier, Greene (1993) investigated the use of citations ("appeals to authority") in student texts. Each instance of explicit reference to the source texts in student texts was coded as one of three functions (pp. 56-57):

1. To use as a source of content: The writer used the work of an author as a source of information
2. To locate a faulty path: The writer attempted to refute a given interpretation by establishing that the line of argument or assertion was mistaken and should be avoided.
3. To support a claim: The writer appealed to an author to support his or her own line of argument.

Greene views the appeal to authorities in relation to making a contribution to a scholarly conversation. Instead of relying heavily on source materials as source of content, writers have an option to use them to present their own perspectives in light of what have been presented by others, interpreting and critiquing them, or using them as support for their argument (p. 68).

Since the problem-based essay task required students to make an original contribution by presenting a proposal, it was expected that students who worked on this task would use more of the third type of appeals whereas those writing a report would resort more to the first type. However, the distributions of the three types were very similar across the two tasks. The majority of appeals to authority were of the first type in both tasks. The mean number of appeals to authority as content was 7.7 for the report task and 6.9 for the problem task. In both tasks, the

support for an argument was much fewer ($M = 2.0$ for report and 3.8 for problem) and faulty path was the least frequent ($M = 0.6$ for report and 1.5 for problem).

In second language research, Campbell (1987, 1990) has used the notions of “foreground” and “background” to define the function of information incorporated from a source text. These notions were taken from the work of Hopper and Thompson (1980). Foreground is “the material which supplies the main points of the discourse” and background is “that part of a discourse which does not immediately and crucially contribute to the speaker’s goal, but which merely assists, amplifies, or comments on it” (Hopper & Thompson, 1980, p. 280). In terms of correspondence between these and Greene’s categories, his first type (content) would be foreground information and his second and third types (faulty path and support) would be background information. Campbell selected thirty college students in three composition courses by a stratified random sampling procedure: 10 less proficient and 10 more proficient ESL students, and 10 L1 English students. The students were first assigned to read the first chapter of an undergraduate anthropology textbook at home. Then they spent an hour in their classes writing a composition. The prompt told them to take the role of an anthropologist planning to observe and report on fraternities and/or sororities in the campus. In their compositions they were to explain which anthropological concepts they planned to consider in their report making reference to the textbook chapter they had read. The explanation needed to be clear enough for a university student with little background knowledge on anthropology. Student texts were parsed into t-units and each separated instance of the use of information from the reading text was categorized according to its function.

The results indicated that the ratio of background and foreground information taken from the source was similar in the texts written by the two groups of ESL students (about half and half). On the other hand, L1 English students’ use of source information was more as background (70%) than as foreground (30%). However, according to the instructor evaluation of holistic quality of source information use, two of the compositions that received the highest score of 4 on a 4-point scale⁴ belonged to more proficient ESL students who used source information primarily as foreground; the remaining two belonged to L1 English students who used source information mostly as background. Thus, it seems that

⁴ 4 = very good, 3 = good, 2 = poor, 1 = unacceptable. The correlation between two instructors was .81. (Campbell, 1987:30)

background use of source information was not necessarily rated higher than foregrounded use. The two ESL students explained the concepts taken from the anthropology text in their own words. In all the three groups, the rhetorical function of more than 85 percent of incorporations from the source text was the definition of concepts from anthropology (Campbell, 1987).

In another study of second language students writing from sources, Lewkowitz (1994) made an impressionistic analysis of function. It appeared that the majority of her students had used information from the two source texts as foreground (introduction of new propositions) rather than as background (support for students' own ideas). They often chose key words in the source texts, and strung them together in their essays as their own ideas. Oftentimes, the words were strung together out of context and without any development making it difficult to follow their argument (p. 213).

3.5 Format of incorporated source information

The format of the source information used in student texts is important in writing assessment because it includes the issue of plagiarism. To date the most thorough investigation into the question of format of source information use has been done by Campbell (1990, 1987). In addition to the type of function discussed in the previous section, each example of source information use in student texts was classified according to the "degree of integration." The following six categories were in the order of the least integrated into the student text to the most integrated. Of these Exact Copies and Near Copies were inappropriate formats, which constituted plagiarism, whereas the other four were appropriate.

Quotations:	direct quotations with the proper punctuating quotation marks
Exact Copies:	direct quotations without the punctuating quotation marks
Near Copies:	similar to Exact Copies with the addition that syntax was rearranged, or synonyms are used for one or two content words
Paraphrases:	involve more syntactic changes of the original text than Near Copies
Summaries:	represent the gist of information from the source text
Original Explanation:	information from the source text is presented through the students' explanation of a theme.

Furthermore, the quality of each example was rated by instructors on a 4-point scale (see Figure 2). In total 188 examples were identified, of which 75 examples received the highest score of '4.' Among these 23 (31%) were original explanations, 14 (19%) were summaries, and 18 (24%) were paraphrases. In Campbell's scheme, these represented the three most integrated formats. Surprisingly, five examples (7%) were near copies and one (1%) was an exact copy. According to Campbell, this indicates that the two instructors did not always notice copying. As for the examples that received the lowest scores, only one received '1,' which was a paraphrase. There were 26 examples that received '2.' Of these the most frequent format was quotations, the least integrated format (ten examples or 37 %). This was followed by near copies and exact copies, the two inappropriate formats (i.e., plagiarism): each six examples or 22%. Campbell does not explain why the appropriate format, quotation, tended to receive the low rating. In terms of function, a total of 20 examples were quotations and of these 16 were backgrounded. However, no other relevant information is offered and it is not clear why half of the quotations received the low score.

-
- 4 True to the student paper and true to the anthropology text. (Appropriate in the paper, provides good support. And correct and appropriate information from the anthropology text.)
 3. One of the above is faulty. The example is either inappropriate in the student paper (plays too major or too minor role in the content of the paper), or it is incorrect/useless information from the anthropology text.
 2. Poor on both points; might be better if left out. Or information is copied from the anthropology text (copied = more than three consecutive words match).
 1. Paper would definitely be better without the example.
-

Figure 2. Four-point scale to evaluate the quality of individual examples of use of information from the background reading text.

Note. From Campbell, 1987, p. 164, Appendix D. The Pearson correlation between two raters was .82. (Campbell, 1987, p. 30)

In addition to the degree of integration, Campbell (1990, 1987) has also analyzed documentation; that is, the reference to the author and/or the title of the source text. Overall, very few attributions were made (see Table 2). Of these the

most of them were on quotations. A few were on paraphrases. In general, ESL students made reference more often than L1 English students.

Table 2

Documentation by Degree of Integration and Student Group (Adapted from Campbell, 1987, Table 2, p. 34)

	Less Proficient ESL	More Proficient ESL	L1 English
Quotation	7/8 ^a	2/2	4/10
Paraphrase	0/4	3/10	1/21
Others	0/38	3 ^b /40	0/55
Total	7/50	8/52	5/86

a. the number of attributed examples/the total number of examples in this category

b. a summary, an exact copy, and a near copy

Campbell offers two possible reasons for the overall lack of documentation. The first reason is that students were not familiar with the convention of documentation in academic writing. However, she argues that a more likely reason is that students felt it unnecessary to cite the source because they had been given only one source text and that they knew their instructors were familiar with the text. It is true that the students were fully aware that their instructors knew that all of them were using the same source text. There seems to be some evidence that at least some students did not feel it necessary to cite the source. Out of the ten quotations used by L1 English students, only four were attributed. The fact that they marked the words of the source by punctuation indicates that they were willing to acknowledge that those words were not their own. Adding citations would have been a simple matter to them. This is exactly what ESL students did with their quotations. However, the same explanation does not apply to the other formats of incorporation from the source. It is not clear whether students did not attribute the source in their paraphrases, summaries, or copies because they were unfamiliar with the convention, they did not feel it necessary under the task condition, or they did not want to diminish their contribution by giving credit to the source author.

Campbell (1990) speculates that if the source texts represent two or more authors with opposing views, students might attribute the sources they use more often (Campbell, 1990, p. 224). As described previously, in the study by Lewkowicz (1994), two source texts were used. All the 32 students in the read

and write condition used ideas from one or both of the source texts. However, only six students “made any attempt to attribute” (p. 211). Moreover, even those who did were not consistent. One student used ideas from a source text which amounted to one third of the essay but the student cited the author only twice (p. 212). Although based on only an impressionistic analysis, Lewkowicz reports that her students used exact copies and near copies extensively. Unlike Campbell, Lewkowicz attributes the extensive copying and the lack of documentation in her students’ essays to the tendency of Hong Kong students to plagiarize. The students she interviewed indicated that they had a very different idea about plagiarism: “the interviewed students themselves said it was easier to copy as in that way they could avoid mistakes. They did not see anything wrong in this as they claimed they were copying from authorities who would inevitably be better than themselves” (Lewkowicz, 1994, p. 212).

4. Conclusions

For a valid use of source-based tasks in writing assessment, it is essential to know two things: in what ways students make use of source materials, and which ways are viewed by instructors as signs of competence or incompetence. However, there are hardly any studies that have investigated these questions. Lewkowicz (1994) has looked at the characteristics of compositions written with two source texts by university freshmen in Hong Kong. But her study did not examine the relationship between the use of source texts and the evaluation by instructors. Although Campbell (1987, 1990) has investigated this issue, her writing task was based on only one source text. There is clearly a need for studies that will look at this problem using multiple source texts. The promising areas in such investigation will be content, integration/organization, function, and format of incorporated source information.

References

- Ackerman, J. (1991). Reading, writing, and knowing: The role of disciplinary knowledge in comprehension and composing. *Research in the Teaching of English, 25*(2), 133-178.
- Bloch, J., & Chi, L. (1995). A comparison of the use of citations in Chinese and English academic discourse. In D. Belcher & G. Braine (Eds.), *Academic writing in a second language: Essays on research and pedagogy* (pp. 3-22). Norwood, NJ: Ablex.

- Breland, H. M., & Jones, R. (1984). Perceptions of writing skills. *Written Communication, 1*, 101-119.
- Bridgeman, B., & Carlson, S. B. (1983). *Survey of academic writing tasks required of graduate and undergraduate foreign students* (Vol. (TOEFL Research Rep. No. 15)). Princeton, NJ: Educational Testing Service.
- Brown, J. D., Hilgers, T., & Marsella, J. (1991). Essay prompts and topics: Minimizing the effect of mean differences. *Written Communication, 8*(4), 533-556.
- Campbell, C. (1987). *Writing with others' words: Using background reading text in academic compositions*. Unpublished PhD Dissertation, UCLA.
- 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: Cambridge University Press.
- Carson, J. E., Carrell, P. L., Silberstein, S., Kroll, B., & Kuehn, P. (1990). Reading-writing relationships in first and second language. *TESOL Quarterly, 24*(2), 245-266.
- Charge, N., & Taylor, L. B. (1997). Recent developments in IELTS. *ELT Journal, 51*(4), 374-380.
- Chiang, S. Y. (1999). Assessing grammatical and textual features in L2 writing samples: the case of French as a foreign language. *The Modern Language Journal, 83*(2), 219-232.
- Currie, P. (1998). Staying out of trouble: Apparent plagiarism and academic survival. *Journal of Second Language Writing, 7*(1), 1-18.
- Durst, R. K. (1987). Cognitive and linguistic demands of analytic writing. *Research in the Teaching of English, 21*(4), 347-376.
- Faigley, L., Cherry, R., Jolliffe, D., & Skinner, A. (1985). *Assessing writers' knowledge and processes of composing*. Norwood, NJ: Ablex.
- Feak, C., & Dobson, B. (1996). Building on the impromptu: A source-based academic writing assessment. *College ESL, 6*(1), 73-84.
- Flower, L. (1990a). Introduction: Studying cognition in context. In L. Flower, V. Stein, H. Ackerman, M. Kantz, K. McCormick, & W. Peck (Eds.), *Reading-to-write: Exploring a cognitive and social process* (pp. 3-32). New York: Oxford University Press.
- Flower, L. (1990b). The role of task representation in reading-to-write. In L. Flower, V. Stein, H. Ackerman, M. Kantz, K. McCormick, & W. Peck (Eds.), *Reading-to-write: Exploring a cognitive and social process* (pp. 35-75). New York: Oxford University Press.

- Flower, L., Stein, V., Ackerman, H., Kantz, M., McCormick, K., & Peck, W. (1990). *Reading-to-write: Exploring a cognitive and social process*. New York: Oxford University Press.
- Freedman, S. W. (1979). How characteristics of student essays influence teachers' evaluations. *Journal of Educational Psychology, 71*, 328-338.
- Ginther, A., & Grant, L. (1996). *A review of the academic needs of native English-speaking college students in the United States* (TOEFL Monograph Series: MS-1). Princeton, NJ: Educational Testing Service.
- Greene, S. (1993). The role of task in the development of academic thinking through reading and writing in a college history course. *Research in the Teaching of English, 27*(1), 46-75.
- Grobe, C. (1981). Syntactic maturity, mechanics, and vocabulary as predictors of quality ratings. *Research in the Teaching of English, 5*(1), 75-85.
- Hale, G., Taylor, C., Bridgeman, B., Carson, J., Kroll, B., & Kantor, R. (1996). *A study of writing tasks assigned in academic degree programs* (TOEFL Research Report 54). Princeton, NJ: Educational Testing Service.
- Hamp-Lyons, L. (1991). Pre-text: Task-related influences on the writer. In L. Hamp-Lyons (Ed.), *Assessing second language writing in academic contexts* (pp. 87-107). Norwood, NJ: Ablex.
- Hamp-Lyons, L., & Kroll, B. (1997). *TOEFL 2000—Writing: Composition, community, and assessment* (TOEFL Monograph Series: RM-96-5). Princeton, NJ: Educational Testing Service.
- Higgins, L. (1993). Reading to argue: Helping students transform source texts. In A. M. Penrose & B. M. Sitko (Eds.), *Hearing ourselves think: Cognitive research in the college writing classroom* (pp. 70-101). New York: Oxford University Press.
- Hilgers, T., & Stitt-Bergh, M. (1993). *Building better bridges from Hawai'i high schools to the University of Hawai'i at Manoa: What Manoa-bound high school graduates' improved essays mean for high school and college teachers* (Manoa Writing Program Report for Teachers #1): University of Hawai'i at Manoa.
- Hoetker, J., & Brossell, G. (1986). A procedure for writing content-fair essay examination topics for large-scale writing assessments. *College Composition and Communication, 37*(3), 328-335.
- Hopper, P. J., & Thompson, S. A. (1980). Transitivity in grammar and discourse. *Language, 56*(2), 251-299.
- Johns, A. (1991). Faculty assessment of ESL student literacy skills: Implications

- for writing assessment. In L. Hamp-Lyons (Ed.), *Assessing second language writing in academic contexts* (pp. 167-179). Norwood, NJ: Ablex.
- Johns, A. M., & Mayes, P. (1990). An analysis of summary protocols of university ESL students. *Applied Linguistics*, 11(3), 253-271.
- Kantz, M. (1990). Promises of coherence, weak content, and strong organization: An analysis of the students' texts. In L. Flower, V. Stein, H. Ackerman, M. Kantz, K. McCormick, & W. Peck (Eds.), *Reading-to-write: Exploring a cognitive and social process* (pp. 76-95). New York: Oxford University Press.
- Kennedy, M. L. (1985). The composing process of college students writing from sources. *Written Communication*, 2(4), 434-456.
- 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: University of Southern California, Southern California Occasional Papers in Linguistics No. 5.
- Kroll, B., & Reid, J. (1994). Guidelines for designing writing prompts: Clarifications, caveats, and cautions. *Journal of Second Language Writing*, 3(3), 231-255.
- Lehner, A., Wolfe-Quintero, K., Tripodi-Segade, G., & Hilgers, T. (in press). *The experiences of bilingual students in writing-intensive courses across the college curriculum* (Manoa Writing Program Technical Report #18): University of Hawai'i at Manoa.
- Leki, I., & Carson, J. (1994). Students' perceptions of EAP writing instruction and writing needs across the disciplines. *TESOL Quarterly*, 28(1), 81-101.
- 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.
- Lewkowitz, J. (1994). Writing from sources: Does source material help or hinder students' performance? In N. Bird & e. al. (Eds.), *Language and Learning: Papers presented at the Annual International Language in Education Conference (Hong Kong, 1993)* (pp. 204-217). (ERIC Document Reproduction Service No. ED 386 050).
- Lewkowitz, J. A. (1997). The Integrated Testing of a Second Language. In C. M. Clapham & D. Corson (Eds.), *Language Testing and Assessment. Encyclopedia of Language and Education, Vol. 7*. Dordrecht: Kluwer Academic Publishers.

- Liu, H.-c. (1997). *Constructing and validating parallel forms of performance-based writing prompts in academic settings*. Unpublished PhD Dissertation, University of Illinois at Urbana-Champaign.
- LoCastro, V., & Masuko, M. (1997, March). *Plagiarism and academic writing of NNS learners*. Paper presented at the Annual Meeting of the Teachers of English to Speakers of Other Languages (31st, Orlando, FL, March 11-15, 1997). (ERIC Document Reproduction Service No. ED 409 724).
- Mathison, M. A. (1996). Writing the critique: A text about a text. *Written Communication, 13*(3), 314-254?
- Norris, J., Brown, J. D., Hudson, T., & Yoshioka, J. (1998). *Designing second language performance assessments* (Technical Report #18). Honolulu: Second Language Curriculum Center, University of Hawaii at Manoa.
- Paltridge, B. (1992). EAP placement testing: An integrated approach. *English for Specific Purposes, 11*, 243-268.
- Pennycook, A. (1996). Borrowing others' words: Text, ownership, memory, and plagiarism. *TESOL Quarterly, 30*(2), 201-230.
- Read, J. (1990). Providing relevant content in an EAP writing test. *English for Specific Purposes, 9*, 109-121.
- Ruth, L., & Murphy, S. (1988). *Designing writing tasks for the assessment of writing*. Norwood, NJ: Ablex.
- Skehan, P. (1996). A framework for the implementation task-based instruction. *Applied Linguistics, 17*(1), 38-62.
- Smith, W. L., Hull, G. A., Land, R. E. J., Moore, M. T., Ball, C., Dunham, D. E., Hickey, L. S., & Ruzich, C. W. (1985). Some effects of varying the structure of a topic on college students' writing. *Written Communication, 2*(1), 73-89.
- Spack, R. (1997). The acquisition of academic literacy in a second language: A longitudinal case study. *Written Communication, 14*(1), 3-62.
- Spivey, N. N. (1983). *Discourse synthesis: Constructing texts in reading and writing*. Unpublished PhD dissertation, The University of Texas at Austin, Newark, DE.
- Spivey, N. N. (1984). *Discourse synthesis: Constructing texts in reading and writing*. Newark, DE: International Reading Association.
- Spivey, N. N. (1990). Transforming texts: Constructive processes in reading and writing. *Written Communication, 7*(2), 256-287.
- Spivey, N. N., & King, J. R. (1989). Readers as writers composing from sources. *Reading Research Quarterly, 24*(1), 7-26.
- Stewart, M., & Grobe, C. (1979). Syntactic maturity, mechanics of writing and

- teachers' quality ratings. *Research in the Teaching of English*, 13, 207-215.
- Sweedler-Brown, C. (1993). ESL essay evaluation: The influence of sentence-level and rhetorical features. *Journal of Second Language Writing*, 2(1), 3-17.
- Vann, R., Lorenz, F., & Meyer, D. (1991). Error gravity: Faculty response to errors in the written discourse of nonnative speakers of English. In L. Hamp-Lyons (Ed.), *Assessing second language writing in academic contexts* (pp. 181-196). Norwood, NJ: Ablex.
- Watanabe, Y. (2001). *Read-to-write tasks for the assessment of second language academic writing skills: Investigating text features and rater reactions*. Unpublished PhD Dissertation, University of Hawai'i.
- Weir, C. (1983). *Identifying the language needs of overseas students in tertiary education in the United Kingdom*. Unpublished PhD Thesis, University of London Institute of Education.
- Weir, C. (1993). *Understanding and Developing Language Tests*. London: Prentice Hall.
- Wesche, M. B. (1987). Second language performance testing: The Ontario Test of ESL as an example. *Language Testing*, 4(1), 28-47.

Department of English
Kanda University of International Studies
1-4-1 Wakaba, Mihama-ku Chiba-shi,
Chiba-ken 262-0014
JAPAN

yu-wata@kanda.kuis.ac.jp