Investigating the quality of source integration in an academic writing exam

Amber Vanderwoude & Liubou Shefarevich
Second Language Testing, Inc., USA

Integrated writing tasks have become the standard for many English language exams. However, do they elicit the type of source use expected in academic settings? To address this question, this study reports on source integration on an operational English exam administered in Japan. Responses of 24 secondary students on two forms of the exam were analyzed and coded for stylistic conventions, language transformation (e.g. paraphrase), and rhetorical functions. While many responses lacked the sophistication and formal conventions of undergraduate- and graduate-level responses analyzed in previous studies, they revealed test takers' emerging ability to engage with sources in a way that reflects the expectations of academic writing.

Keywords: integrated tasks, writing assessment, source use, EFL writing, EFL assessment

Introduction

Writing skills are often an integral part of language proficiency both for native speakers of English and learners of English as a Foreign (EFL) or Second Language (ESL). To participate in English language education either at home or abroad, EFL and ESL students must be able to produce writing based on a variety of sources such as research studies, articles, discussion boards, and more.

At universities in the United States (US), integrated writing skills feature prominently in the curricula of English writing courses for first-year students. Our review of course descriptions from several universities shows that the goal of such courses is to support

Email addresses for correspondence: avanderwoude@2lti.com, lshefarevich@2lti.com

students in carrying out academic research, which involves identifying relevant sources, selecting relevant information, and incorporating it into their own writing by means of summary, paraphrase, and synthesis. Students also acquire the procedural knowledge to perform these tasks responsibly by properly citing the sources (Harvard University, 2023; Massachusetts Institute of Technology, 2014; New York University (n.d.); University of Illinois at Urbana-Champaign, 2019–2020; University of Southern California (n.d.)). However, to gain admission into US universities, many international students need to pass standardized English proficiency tests that may already include integrated writing tasks, requiring students to display integrated writing skills before they have even entered university.

Many universities in Japan also require students to pass an English proficiency test for admission and to take English academic writing courses in their first year (Moore, 2015; Shiratori, 2019). These tests often reflect the national school curriculum and affect the choice of constructs being taught to high school students. Moore (2015) investigated the connection between teaching/learning and assessment in Japan, comparing the tasks on English tests for university admission and the writing constructs in first-year English academic writing courses at 51 private and public universities. The writing tasks in the admission tests typically involved formats such as translation, word reordering, and lexical gap-filling. Free, or impromptu, writing tasks were included in writing tests at only 11 out of 39 public universities and at three out of 12 private universities. In contrast, surveys conducted with lecturers at these universities showed that students in first-year English academic writing and subject courses were expected to write essays, summaries, and reports. Citing sources, synthesizing, evaluating, and expressing one's own views were also found to be important language functions. Thus, Moore's analysis suggests that there is a gap between the skills assessed on English entrance tests in Japan and the skills needed to successfully participate in university-level English courses at home or abroad.

Reflecting the importance of this skill, integrated writing tasks have been added to many other tests of English language proficiency, including those used for university admissions in the US, such as the TOEFL iBT (Knoch & Sitajalabhorn, 2013; Plakans & Gebril, 2012). If integrated writing tasks in language tests can elicit the type of writing

that is expected in academic contexts, they may inform and motivate teaching and learning and serve as more accurate tools for assessing test takers' preparedness for these contexts (Cumming, 2013; Cumming et al., 2006; Moore, 2015). There is research suggesting that writing produced on independent writing tasks and on integrated writing tasks is qualitatively different (Cumming et al., 2006). For example, Cumming et al. (2006) compared discourse features of responses by adult ESL students on prototype TOEFL integrated listening-reading-writing tasks to existing TOEFL independent writing tasks. On integrated tasks, writers exhibited more lexical variety and higher syntactic complexity. They constructed extended written arguments and supported them with information from the sources by means of paraphrase, summary, and verbatim rendering. According to this analysis, integrated tasks may indeed elicit the type of writing that is expected in the university classroom. However, more research is needed that focuses on writing performance of secondary students, for whom integrated writing skills are of high importance when considering entrance into, and academic success in, universities. Our study addresses this gap by investigating source integration produced by secondary students in Japan on an operational English proficiency exam.

Review of existing research

Integrated writing skills have previously been analyzed using samples of writing from language learners in various contexts, including high school, undergraduate, and graduate settings (Gebril & Plakans, 2009; Hirvela & Du, 2013; Iwata, 2012; Kato, 2018; Lee et al., 2018; Petrić, 2007). Such studies can help provide the methodology for analyzing writing produced on integrated test tasks to determine how closely it resembles the writing produced in the target language use domain. They also show that integrated writing may present a challenge to L2 writers even at advanced levels of study.

To successfully complete integrated writing tasks, language learners must have paraphrasing and summarizing skills in addition to the skills of reading (or listening) and writing. Hirvela & Du (2013) define paraphrase as a process of recasting individual sentences, creating a combination of original language and grammatical structures from the source text with new words and grammatical structures. They point out that

"paraphrasing, because of what it asks students to do with the original source material, provides insight into how well students read (since comprehension is the first step toward paraphrasing) as well as write" (p. 88). Using think-aloud protocols and text-based interviews, they analyzed paraphrasing practices by two adult ESL students from China. It was found that, while student writers may have mastered the mechanics of paraphrasing through decontextualized classroom activities, they may not realize how to use this technique for constructing an effective argument and for integrating information from the source(s) with their own knowledge. As a result, they may often resort to inserting excerpts from the source(s) into their own writing simply to demonstrate superficial familiarity with the source(s), either using direct quotes or engaging in "patchwriting," a term first used by Howard (1992), which involves copying phrases or sentences from a source almost verbatim and only making minimal surface-level modifications.

Paraphrasing is a precursor to summary writing, as shown in research by Cumming et al. (2006). In their study of writer performance on prototype integrated-skills tasks for TOEFL, paraphrasing was more commonly used by midrange-level writers and summarizing by higher proficiency writers. Summary writing is a more complex skill and involves condensing the most prominent idea(s) of a source text to create a shorter version. It can also involve generalizing on several key points of the source text. In any case, constructing a summary involves both reading (or listening) and writing (Hedgcock & Ferris, 2009; Hirvela & Du, 2013).

In addition to paraphrasing and summarizing sources efficiently, student writers are also expected to use them for a variety of rhetorical functions, such as attribution, exemplification, and others, to establish connections between their own knowledge and the overall field of knowledge represented in the sources. Studies by Petrić (2007) and Lee et al. (2018) provided a framework for investigating the quality of source integration in our study. Building upon different typologies, Petrić (2007) conducted quantitative and qualitative analysis of rhetorical functions in sixteen Master's theses written by L2 graduate students. The purpose of the analysis was to identify the differences in rhetorical functions of citations between low- and high-rated theses. It was found that overall

citations were used more frequently and with higher density in high-rated than in low-rated theses, and attribution was the most commonly used rhetorical function type in both groups. Petrić (2007) points out that "this may be characteristic of student writing in general, since this citation function helps writers display their knowledge of the topic" (p. 247). However, while in low-rated theses citations were used *mostly* for attribution, in high-rated theses sources were represented through a wider variety of functions such as evaluation, establishing links between sources, and establishing connections between students' own research and sources.

Lee et al. (2018) analyzed citation practices in 100 high-rated research papers by L2 students in a first-year writing course at a US university. They examined the use of surface forms, rhetorical functions, and writer stance. In terms of surface forms, they classified citations into four categories: direct quotation, block quotation (a string of 40 or more words from a source), summary (information from one source), and generalization (synthesized information from two or more sources). The analysis shows that citations were used most frequently and "overwhelmingly in the form of summary" (p. 5), followed by a small percentage of direct quotations. Generalizations were rarely observed. These data may suggest that "the L2 [undergraduate] student writers have considerable difficulty in synthesizing multiple sources in advancing their argument and supporting their claims" (p. 6). In their analysis of rhetorical functions, they applied the typology used by Petrić (2007). Similar to Petrić's study (2007), attribution was the most frequently used function. A small percentage of citations were used for evaluation, exemplification, and establishing links between sources.

Finally, while it is important for language learners to develop and demonstrate the ability to integrate outside sources into their own writing, in university-level writing it is also important to follow the established procedures for referencing the sources. A primary reason may be because plagiarism remains a concern at the undergraduate and graduate levels. However, some research points to the fact that copying strings of words without properly acknowledging the source(s) could be a result of limited language ability and/or lack of formalized training (Chan & Yamashita, 2022; Cumming et al., 2006; Gebril & Plakans, 2009; Shi, 2004). As Gebril & Plakans (2009) show, even when language

learners are aware of plagiarism being an issue, they might still resort to inappropriate textual borrowing because integrated writing might be new to them or because they lack the procedural knowledge of acknowledging sources. Integrated tasks on language tests might be a way to establish the students' level of citation skills and to inform instruction in this area as well.

Although previous research offers analysis of different aspects of integrated writing in various contexts, we are not aware of any studies that have analyzed multiple features of integrated writing by secondary students in operational language proficiency tests. To address this gap, we present a small-scale qualitative study into characteristics of source use by secondary students in Japan on an academic English writing proficiency exam, Second Language Testing Inc.'s (SLTI) *Written Communication Exam for Academia*. Analyzing a range of subskills, such as acknowledging sources, paraphrasing and summarizing, and rhetorical functions, can provide a more complete view of a writer's overall ability to integrate sources into their writing (Lee et al., 2018). Doing so within the context of an operational exam can determine whether the exam elicits key subskills associated with source integration. Thus, the aim of our analysis was to discover what kind of source integration was being elicited by the integrated writing task on our exam. To do this, we broke down our primary research question into three parts:

RQ: How are test takers integrating the sources into their writing?

a) What stylistic conventions are test takers using?
b) How are test takers transforming the language?
c) How are test takers synthesizing ideas from the sources with their own ideas?

Methods

Participants

The participants consisted of 24 secondary students in Japan. Twenty-two of the students participated in two online administrations of SLTI's *Written Communication Exam for Academia*, completing two forms of the exam. The exam is scored according to the levels of the Common European Framework of Reference for Languages (CEFR) (Council of

Europe, 2020), and the participants were selected because they received an overall score in the B1 or B2 range on Form 1 of the exam. The two additional participants participated in only one administration. However, they were included because they were the only test takers to receive an overall score of High B2, so we expected their writing to best reflect the type of writing we were hoping the exam would elicit.

Materials

SLTI's *Written Communication Exam for Academia* is an online exam with three writing tasks. The final task is an integrated reading and writing task that was designed to measure, among other constructs, a test taker's ability to integrate sources into their writing. To determine how well the task elicits this ability and, by extension, whether it is appropriate to interpret the result of the exam as an indicator of academic writing ability (Carr, 2011, p. 20), we conducted a qualitative analysis of responses to the integrated task.

The integrated task

The integrated reading and writing task instructs test takers to read two B2-level passages that present differing views on a topic and then write an opinion-based, essay-like response to a question about the topic. On each form of the exam, the integrated task covers a general academic topic, for instance, charging entrance fees at national parks. Topics were deemed appropriate based on our development team's review of multiple English language writing textbooks used by the target audience. Reading passages for the task are presented as articles, blog posts, or discussion forum posts and include features that provide opportunities for citations: fictional titles, authors, and dates. To ensure the level of the reading passages was appropriate for the task, they were written and reviewed by experienced item writers familiar with the CEFR descriptors and the corresponding English Vocabulary Profile (Cambridge, 2023). As recommended by Plakans and Gebril (2012), the task instructions explicitly "refer to using the sources and make it clear that the writers [should take] a position on the topic" (p. 23).

The two forms covered different topics, were presented in different formats, and included slightly different instructions. The passages on Form 1 were presented as discussion forum posts, and the instructions required test takers to include information from both

passages, whereas the passages on Form 2 were presented as online articles, and the instructions did not explicitly require test takers to refer to both passages. Form 1 passages, along with both iterations of the task instructions, can be found in Appendices A and B.

The rubric

The rating rubric, which assigns levels based on the CEFR descriptors, explicitly focuses on source use and integration because they are an important aspect of academic writing and one of the constructs the exam aims to assess. According to the CEFR descriptors for written reports and essays, at the higher end of the B1 level, language users "Can summarise, report and give their opinion about accumulated factual information on familiar routine and non-routine matters within their field with some confidence" (p. 68). At the lower end of B2, they "Can synthesise information and arguments from a number of sources" (p. 68). Within our rubric, varying degrees of integration are expected at the High B1, Low B2, and High B2 levels, as shown in Table 1.

Table 1. Source integration in the rubric

	Quality of integration
High B2	Evaluation and seamless integration of source(s)
Low B2	Synthesized referencing of source(s)
High B1	Mentioning/referencing of source(s)

While language users at the High B1 level should be able to summarize, given the B2 level of the reading passages, simply mentioning the source material in a way that demonstrates minimal understanding was deemed adequate for High B1. It is also important to note that, although the task is set up to encourage formal citation, it is not required by the instructions and therefore not included in the rubric. The construct in our exam focuses on the integration of ideas rather than the formal stylistic conventions used when citing sources.

Procedure

We compiled responses from two administrations of the exam, which took place approximately six months apart. The exams were administered online and double rated by experienced raters. The data sample consisted of 46 essay responses from 24 test

takers: two responses (Form 1 and Form 2) from the 22 who participated in both administrations, and one response each from the additional two test takers.

The 46 test taker responses were coded using categories for three aspects of integration: *stylistic conventions*, *language transformation*, and *rhetorical purpose.* Because of time limitations and the detailed analysis required for qualitative coding of the responses, the coding was distributed among four coders, three of whom were involved in the initial development of the exam and all of whom had extensive experience developing items for and examining responses to similar academic writing exams.

Following recommendations set out by O'Conner and Joffe (2020), we started by coding eight responses (17.4%) together to familiarize ourselves with the codes, adjust definitions, and discuss potential challenges. The rest of the coding took place in three stages, one for each category of integration. Responses were first coded for *stylistic conventions:* six responses (13%) were double coded and then reviewed to resolve discrepancies as a team before the remaining 32 (69.6%) were divided and single coded. The same method was used to code for *language transformation* and *rhetorical purpose.* The rate of agreement prior to review can be found in Table 2.

While there is a lack of consensus regarding the benefits of calculating intercoder reliability, particularly using percentages, doing so can shed some light on how consistently the categories were applied by the coders (O'Conner & Joffe, 2020). Because the *stylistic conventions* category seemed to be more straightforward than the *language transformation* and *rhetorical purpose* categories, we spent less time discussing it before and during the coding process. This may explain why the agreement was low compared to the more linguistically complex category of *rhetorical purpose.*

Table 2. Intercoder agreement

	Pair 1	Pair 2	Average
Stylistic Conventions	67%	80%	73%
Language Transformation	75%	60%	67.5%
Rhetorical Purpose	83%	75%	79%

Although every instance of a particular category, such as quotation marks or evaluation, was tagged in the responses, each code was only counted once per test taker response. Doing so allowed us to calculate how many test takers demonstrated the ability to use a

specific subskill, rather than which subskills were used most frequently, but potentially by only a small percentage of test takers.

Analysis

We conducted a qualitative analysis of all 46 responses, identifying examples of the types of source integration specified in our research question: stylistic conventions, language transformation, and synthesis of ideas (rhetorical purpose).

Stylistic conventions

Although the use of formal citation conventions was neither required nor considered in rating, "the appropriate conventions of documenting and citing sources" (Harvard University, 2023) are expected in university level academic writing, as indicated in writing course descriptions at US institutions such Harvard, the Massachusetts Institute of Technology, and more, and in surveys conducted at Japanese universities. Thus, investigating which stylistic conventions test takers used in their responses was one way of determining how closely our task resembles target language use tasks and can therefore be considered "authentic" (Carr, 2011, p. 20).

To create our list of stylistic conventions, we used a combination of patterns observed in test taker responses and standard citation conventions suggested in style guides like the *Publication Manual of the American Psychological Association* (APA, 2020) and the *Chicago Manual of Style* (CMoS) (The University of Chicago, 2017), for example, the use of author names, titles, dates, and more. Test taker patterns that fell outside of the standard citation conventions led us to include categories like (d) generic reference, where test takers clearly acknowledged that information had been borrowed from an outside source but did not use the author's name or a title in their reference, for example, "According to the article, children treated the visual dogs in the game like live ones." Another such category was (e) generic reference (secondary), where test takers referred to studies mentioned in the source passages, but also without including the title or author's name. The final non-standard category was (f) missing citation. We included this category to recognize test takers' ability to incorporate ideas from a source even if they

did not acknowledge the source. A complete list of the categories along with examples can be found in Table 3.

Table 3. Stylistic convention categories

Category	Description	Example
(a) Quotation Marks	Uses “”	“People who own pets have lower blood pressure [...] than those who don't.” [TT1 Form 2, B2]
(b) Reference to author	Includes the name of the relevant author from the source	Smith thinks that we should charge them... [TT23 Form 1, B2]
(c) Reference to title	Includes the title of the relevant article or post	Although the article “Human's New Best Friend” by Y. Kim said that artificial pets provide similar sense of responsibility... [TT16 Form 2, B2]
(d) Generic reference	Makes a generic reference to the article, post, author, some people, etc.	According to the article, children treated the visual dogs in the game like live ones... [TT22 Form 2, B2]
(e) Generic reference (secondary)	Refers to a study mentioned in the source passage (with no mention of the author/title of the source passage)	Some studies prove that the pet owners have lower blood pressure... [TT 8 Form 2, B1] For example, the article says that studies show that pet owners are less likely to experience depression than people without pets... [TT7 Form 2, B1]
(f) Missing citation	Ideas from a passage are used without any acknowledgement of the source	First, entrance fees provide essential funding for the park. [TT9 Form 1, B1]

Language transformation

The second aspect of source integration that we analyzed was language transformation. In both first- and second-language academic writing in English, writers are expected to effectively quote, paraphrase, and/or summarize source material to demonstrate understanding of the material and to avoid plagiarism (Harvard University, 2023; Massachusetts Institute of Technology, 2014; New York University (n.d.)). Thus, we began with these categories and once again adjusted our list based on patterns that emerged in the test takers’ responses.

We divided the concept of quoting into two categories: (a) direct quote and (b) copying. Both categories involved no transformation of the language. However, direct quote

acknowledged the source in one of the ways mentioned in Table 3(b–e), while copying did not, i.e. the citation was missing (see Table 3(f)). Again, we included both categories to recognize test takers' attempts at using sources even if they did not appropriately cite the source(s), an oversight that would typically constitute plagiarism by most academic standards (APA, 2020).

We defined (e) summary as either the main idea or a combination of multiple key ideas from a source and (c) paraphrase as the transformation of a single key idea or detail from a source (Cumming, et al., 2006; Hedgcock & Ferris, 2009; Hirvela & Du, 2013).

As others have before us (Hirvela & Du, 2013; Howard, 1992; Yamada, 2003), we ran into the issue of whether the attempts to paraphrase were successful or whether they would likely be flagged as plagiarism in an English-medium university classroom. To distinguish between the two, we added the category (d) minimal paraphrasing and defined it as source use that demonstrated the transformation of a few words, perhaps to make it fit grammatically into a sentence or to connect related ideas that came from one area of a passage. We viewed it much like Howard's concept of "patchwriting," where the mere substitution or deletion of words or changes in grammar may result in a "too close paraphrase" (1992, pp. 236–237), but one which teachers or assessors may choose to recognize, rather than penalize. Each of the definitions and examples used when coding can be found in Table 4.

Table 4. Language transformation categories

Type	Definition	Example
(a) Direct quote	Word-for-word copy of key ideas, with a reference but with or without ""	According to the article, *dog owners even get sick less frequently than people without dogs. Although we tend to associate germs with illness, being exposed to the diverse mix of bacteria and other germs that dogs carry can actually equip the immune system to fight illness*. [TT3 Form 2, B1]
(b) Copying	Word-for-word copy of key ideas with or without "". No acknowledgement of the source	I know that *high entrance fees prevent many citizens from enjoying parks*. [TT7 Form 1, B1]
(c) Paraphrase	Rephrasing a single idea from a passage	In addition, people who have a pet *can improve social health as well*. They communicate a lot and they make up good *relationship* with neighbor. [TT5 Form 2, B1]
(d) Minimal paraphrasing	Mostly copied but changes a few words; would likely not be accepted by a college writing instructor	*Dog owners*, for example, *get sick less frequently than people without dogs*. This is beacuse our *immune system* will be *exposed to the diverse mix of bacteria and other germs that dogs carry*. [TT6 Form 2, B1]
(e) Summary	Main idea or a combination of multiple key ideas from a source	S.Kim said the park should not chahrge the fees. She explained that the *beauty* of the national park is a *community resource*, so it need to be *available to all* families. When the park charge the fees, a family with *low-income* cannot visit. [TT22 Form 1, B2]

Note. The reading passages for Form 1 are included in Appendix A. Key language that was directly copied from a passage is *italicized* above and in the Results section.

Rhetorical purpose

The final aspect of source integration that we analyzed was how test takers synthesized information from the sources with their own ideas. To do this, we attempted to identify the rhetorical purpose of an instance of source use to understand how ideas borrowed from the sources functioned within the test taker's response and related to the test taker's ideas, for example, whether information from a source was used as a reason to support a test taker's opinion or whether a test taker's attitude was revealed through the way in which they evaluated information from a source. We drew upon the frameworks used by Lee et al. (2018) in their investigation into the citation practices of first-year L2 university students and Petrić's (2007) analysis of the rhetorical function of citations in Master's theses, starting with the categories used in these analyses and then adjusting the

categories to fit patterns that emerged in the writing of our test takers, who were secondary students. The seven categories we ultimately decided upon are included in Table 5, along with a definition and example of each. Some notable adjustments we made to the existing frameworks included relabeling attribution as reporting, since the information reported was not always attributed to the source(s), and adding categories for reasons, counterarguments, and misrepresentation.

Table 5. Rhetorical purpose of source use

Type	Definition	Example
(a) Reporting	Stating facts from a source to demonstrate knowledge (doesn't serve any other rhetorical purpose beyond conveying facts, may or may not acknowledge the source)	*In 2016*, *1.5 million* people have visted *Plitivse Lakes National Park* and *5 million* people have visited *Yosemite National Park*.And for National Park`s security the goverment gives *$2.5 billion* each year. [TT6 Form 1, B1]
(b) Providing a reason	To support an opinion or position	As a second reason, the article says, "People who own pets *have lower blood pressure, less stress, and are less likely to suffer from depression*…" [TT1 Form 2, B2]
(c) Exemplification	To provide "specific evidence supporting a more general claim" (Petrić, 2007)	First, people with living pet are likely to live a healthy life.*Dog owners*, for example, *get sick less frequently than people without dogs*. [TT6 Form 2, B1]
(d) Evaluation	To evaluate source(s) using evaluative language (such as *agree, support, good, bad*)	I strongly support Smith's idea that national parks charge *entrance fees* Can be used for making good environment is the park for visiter. Also I can not agree with Kim's idea that visiters do not need to pay *fees* because the parks has already received *funding* from *goverment*. [TT4 Form 1, B1]
(e) Establishing links between sources	To point to links between different sources	Though having no *entrance fees* will make it easier for *families with limited budgets* to come to the parks, *charging an entrance fee can also reduce crowding which has a negative* effect on the environment. [TT15 Form 1, B2]
(f) Counter-argument	Contrasting the ideas from a source with the test taker's own ideas to demonstrate the logic of the test taker's idea	Although the article "Human's New Best Friend" by Y. Kim said that artificial pets provide similar *sense of responsibility, companionship and acceptance*, I think with live animal with emotions, you feel more connection. Robots can be programmed or be made again and again. Live animals has a life and can't be made again. This leaves us with responsibility and speciality to the owner. [TT16 Form 2, B2]
(g) Misrepresentation of source	Misrepresenting the facts presented in a source	As stated in the article, interacting with living creatures can activate the brain and prevent not only *depression* but also dementia. [TT 1 Form 2, B2]

Note. The reading passages for Form 1 are included in Appendix A. Key language that was directly copied from a passage is *italicized* above and in the Results section.

Results

Because the reading passages and writing prompts on Forms 1 and 2 addressed different topics, were presented in different formats, and included slightly different instructions, we analyzed the results of each form separately. We did this to identify potential influence the content or format of the sources might have had on how test takers integrated them into their responses. For each form and each category—*stylistic conventions, language transformation,* and *rhetorical purpose*—we tallied the total number of B1 and B2 responses that used a particular convention. For Form 1, we reviewed the responses of 23 test takers: 16 had received a B1 on the task and 7 a B2. For Form 2, we also reviewed 23 responses: 13 had received a B1 and 10 a B2. Our findings are broken down by category below.

Stylistic conventions

Over the two administrations of the test, 91.7% (22) of the 24 test takers acknowledged the sources at least once in their response to either Form 1 or Form 2, with some using multiple citation conventions throughout a response. However, different patterns in the use of stylistic conventions emerged on each form, as shown in Figures 1 and 2. Reference to author was the predominant convention on Form 1, at nearly 70%, and was incorporated using a variety of reporting verbs:

> (1) S.Kim said the park should not chahrge the fees. She explained that … [TT22 Form 1, B2]
>
> (2) Smith thinks that we should charge them… [TT23 Form 1, B2]

Generic references (Example 3) and generic references (secondary) (Examples 4–5) were the most frequently used conventions on Form 2, with some test takers replacing their use of author names with more generic references, like “article” or “studies.” Again, there was some variety in the verbs used to cite sources:

> (3) According to the article,children treated the visual dogs in the game like live ones… [TT22 Form 2, B2]
>
> (4) Some studies prove that the pet owners have lower blood pressure… [TT 8 Form 2, B1]
>
> (5) For example, the article says that studies show that pet owners are less likely to experience depression than people without pets… [TT7 Form 2, B1]

None of the test takers included dates, and only one included quotation marks, although both are common requirements of citing sources in academic writing. However, test takers clearly attempted to use some standard stylistic conventions, despite not being explicitly asked to do so in the instructions. Although they were not consistently referring to an author or a title, as would be required in formal academic writing, they were consistently acknowledging when an idea was not their own. Moreover, many test takers seemed to be reading and writing carefully enough to attempt secondary citations, exhibiting some awareness of this more cognitively and linguistically complex convention.

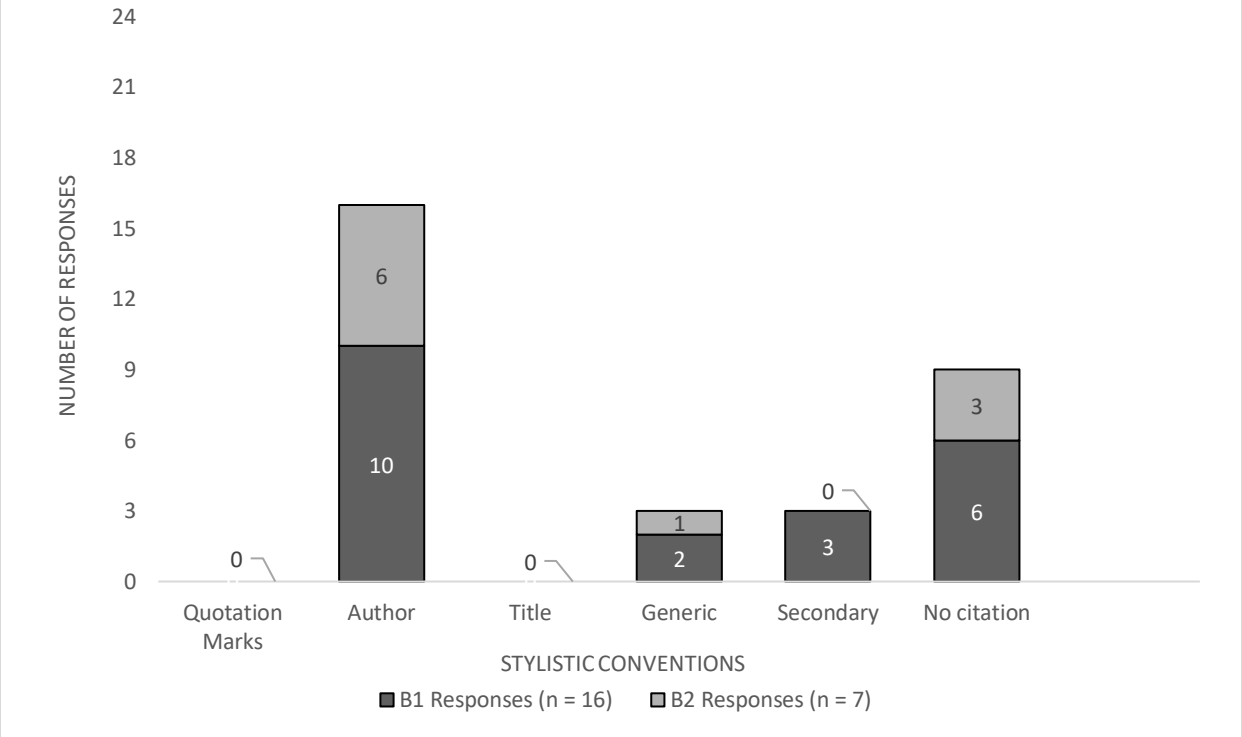

Figure 1. Stylistic conventions used in B1 and B2 responses on Form 1

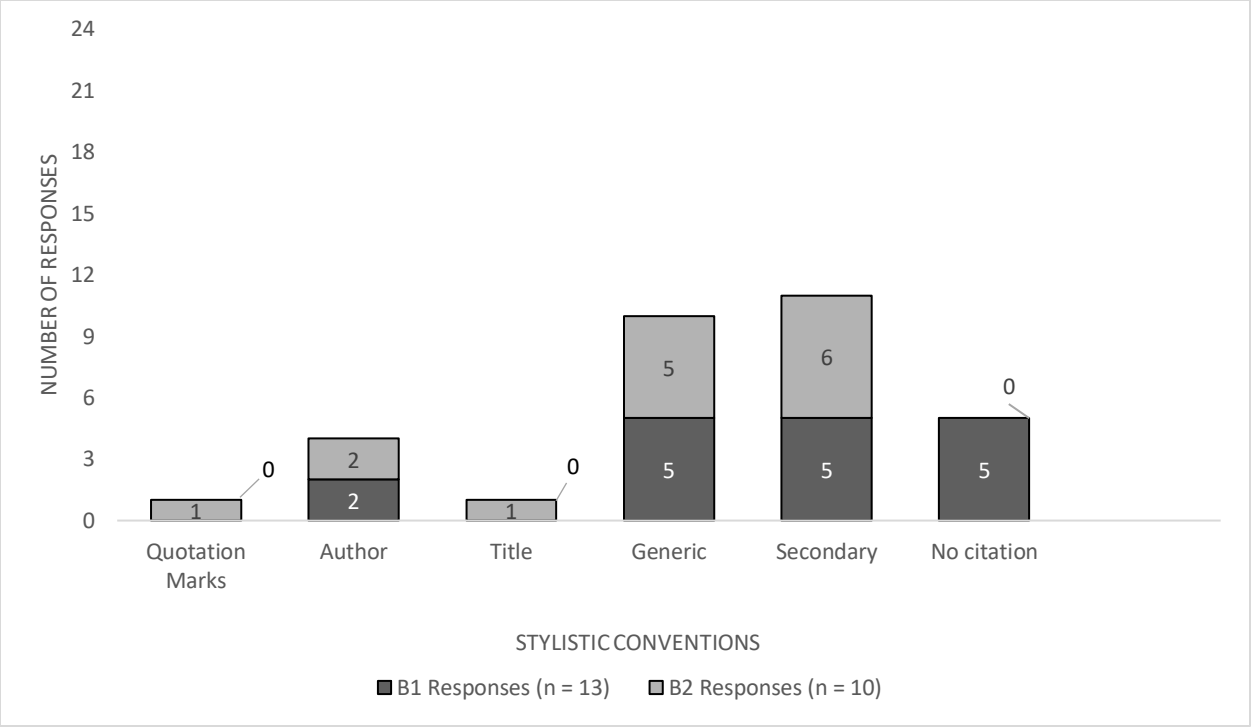

Figure 2. Stylistic conventions used in B1 and B2 responses on Form 2

Language transformation

The ways in which test takers transformed the language can be seen in Figures 3 and 4. In both administrations of the test, nearly half demonstrated the ability to successfully transform the language through either paraphrase or summary, though summaries only appeared in the responses on Form 1. In Example 6, the test taker rephrased this single key point from the source: "Having a pet can improve social health as well. A 2015 survey found that pet ownership not only facilitates interactions between strangers, it can also lead to more meaningful relationships."[2]

> (6) In addition, people who have *a pet can improve social health as well*. They communicate a lot and they make up good *relationship* with neighbor. [TT5 Form 2, B1]

On the other hand, in Example 7, the test taker combined the author's overall position with two supporting reasons, condensing information found throughout the source (Appendix A, Parks for All).

[2] Language that is directly copied from a source is represented in *italics* in the examples for the Language Transformation and Rhetorical Purpose sections.

> (7) S.Kim said the park should not chahrge the fees. She explained that the *beauty* of the national *park is a community resource,* so it need to *be available to all* families. When the park charge the fees, a family with *low-income* cannot visit. [TT22 Form 1, B2]

Of the 11 test takers who incorporated summaries into their responses, nine organized their responses such that information from the sources was often presented separately from their own opinions, rather than being woven together in a way that better demonstrated the ability to "integrate their own ideas with those of others" (University of Southern California, n.d.).

> (8) First, I will talk about my two fellow classmate's opinions. Smith thinks that… [TT23 Form 1, B2]

In the responses on Form 2, test takers did a better job of weaving information from the sources with their own language and ideas, but they often did it through direct copies or quotes (56.6%).

> (9) Secondly, they would develop children's responsibility. According to the article, *children treated the* visual *dogs in the game like live ones, investing time in* thier *care and protecting them from digital dangers*. It is a good trend to children. [TT22 Form 2, B2]

Across both forms, there was no change in the overall number of test takers (7) who transformed the language with limited success, using a minimal paraphrase. For instance, in Example 10, the test taker communicated one key idea from the passage but only transformed some functional language like "for example." The rest was directly copied from the source.

> (10) *Dog owners,* for example*, get sick less frequently than people without dogs*. This is beacuse our *immune system* will be *exposed to the diverse mix of bacteria and other germs that dogs carry*. *[TT6 Form 2, B1]*

While this level of transformation seems to closely align with the B1 level of paraphrasing ability: "paraphrase short passages in a simple fashion, using the original text wording and ordering" (CEFR, 2020, p. 101), it was utilized by test takers who received both B1 and B2 ratings on their responses. However, as with stylistic conventions, some test takers used multiple approaches to language transformation or used the same approach more than once in a response, providing a more complete picture of their abilities.

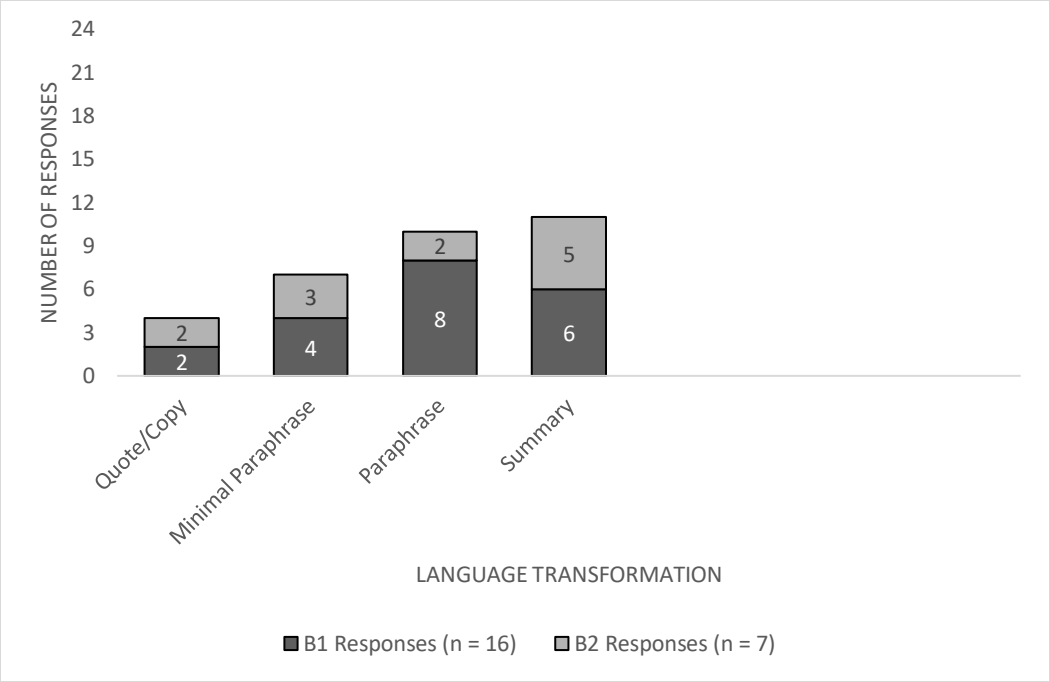

Figure 3. Language transformation in B1 and B2 responses on Form 1

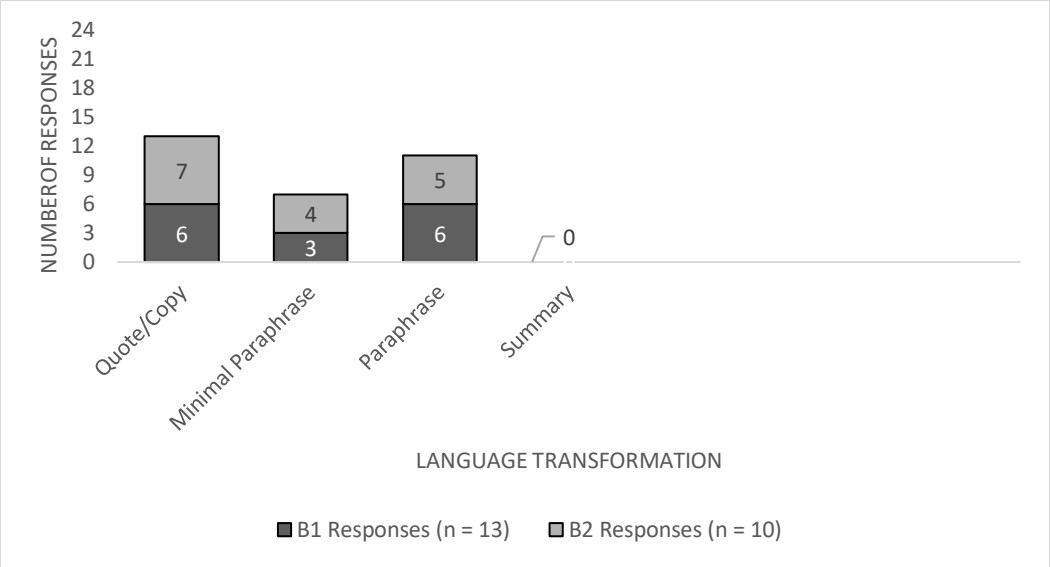

Figure 4. Language transformation in B1 and B2 responses on Form 2

Rhetorical purpose

Finally, test takers used a variety of rhetorical purposes to synthesize the sources with their own ideas, as can be seen in Figures 5 and 6. On both forms, test takers mostly used ideas from the sources as reasons or examples to support their arguments, which was not

surprising given that "reasons and examples" were requested in the instructions. Although the test takers were instructed to use their own reasons and examples, they were not forbidden from using the ideas in the sources as well, as the test takers in the following examples did.

> (11) As a second reason, the article says, "People who own pets *have lower blood pressure, less stress, and are less likely to* suffer from *depression*…" [TT1 Form 2, B2]
>
> (12) First, people with living pet are likely to live a healthy life.*Dog owners*, for example, *get sick less frequently than people without dogs*. [TT6 Form 2, B1]

However, on Form 1 there was greater variety in the rhetorical purposes used. Of note are establishing links between sources and building of counterarguments, even in B1 responses. In Example 13, the test taker links the idea of limited park access from one source and reduced crowding from the other source (Appendix A) to present both sides of the issue of charging fees.

> (13) Though having no entrance fees will make it easier for families with limited budgets to come to the parks, charging an entrance fee can also reduce crowding which has a negative effect on the environment [TT15 Form 1, B2]

In Example 14, the test taker uses the author's position in favor of artificial pets to build a logical argument against them.

> (14) Although the article "Human's New Best Friend" by Y. Kim said that artificial pets provide similar *sense of responsibility, companionship and acceptance*, I think with live animal with emotions, you feel more connection. Robots can be programmed or be made again and again. Live animals has a life and can't be made again. This leaves us with responsibility and speciality to the owner. [TT16 Form 2, B2]

Finally, there were very few instances of reporting, indicating that at all levels, most test takers were attempting to do more than just convey facts from the sources; they were consistently connecting information from the sources to other ideas throughout their responses.

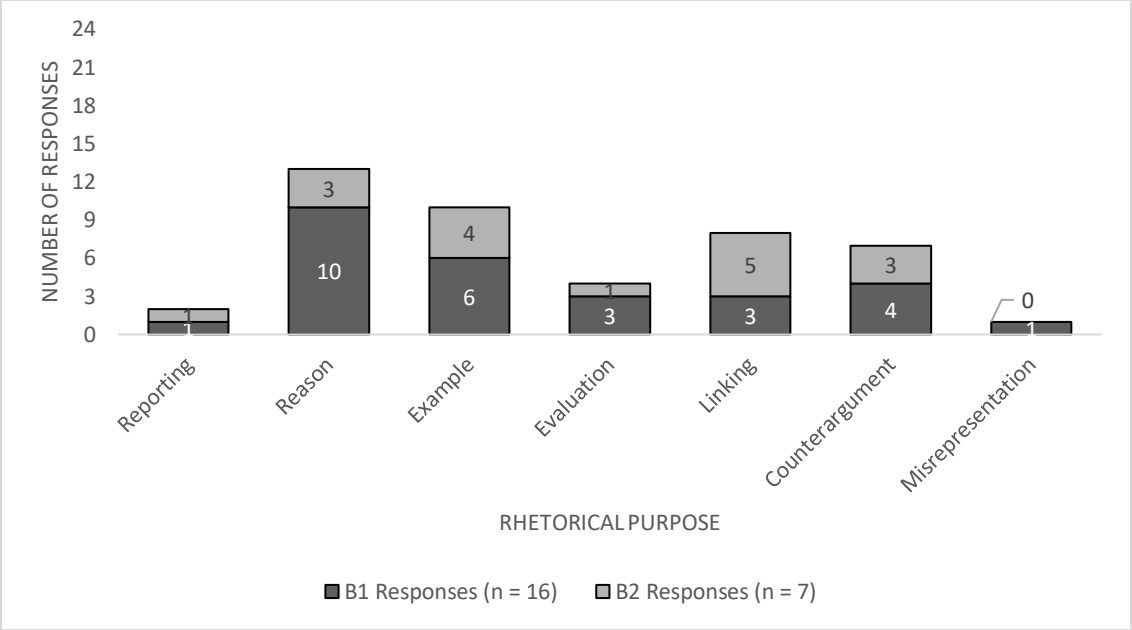

Figure 5. Rhetorical purposes used in B1 and B2 responses on Form 1

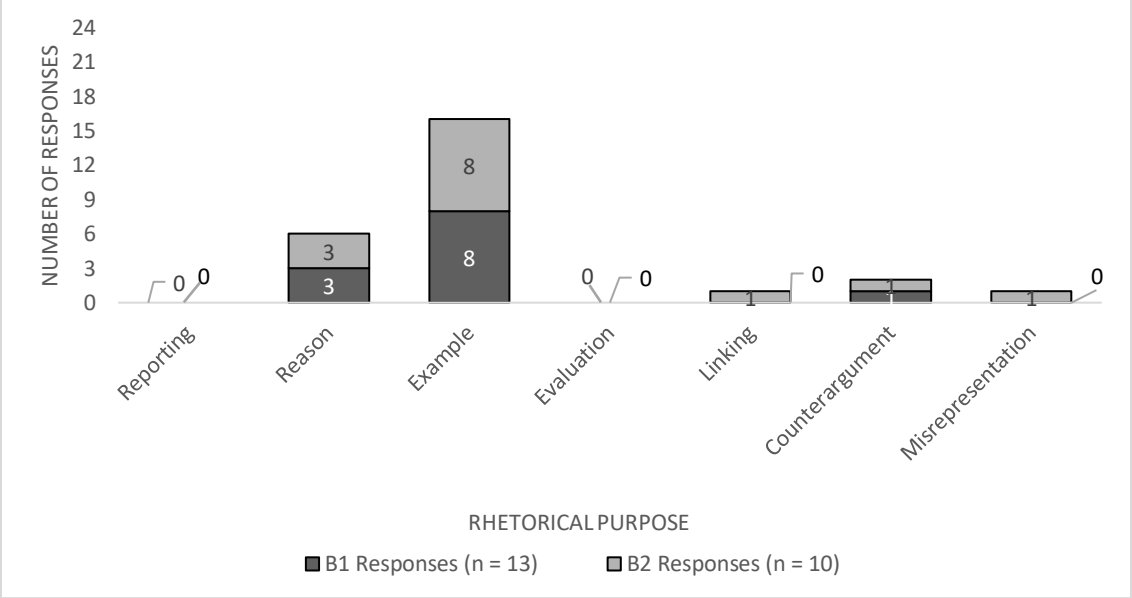

Figure 6. Rhetorical purposes used in B1 and B2 responses on Form 2

Discussion

The purpose of this study was to discover what kind of source integration was being elicited by the final writing task on our exam by analyzing stylistic conventions, language transformation, and rhetorical purposes. Overall, the 24 test takers displayed a range of

skills associated with integrating sources into writing, including skills that were not explicitly requested in the instructions, like citing the sources. Many test takers attempted to transform the language of the sources, albeit not always successfully. They also employed a variety of rhetorical purposes, including some complex ones, like counterargument. Distinct patterns emerged in the two administrations of the exam, which highlighted the possible impacts that prompt instructions, formatting, and source content might have had on the writing task. Key findings corresponding to the three research questions addressed in the study are discussed below.

Stylistic conventions

As stated in the Methods, "the appropriate conventions of documenting and citing sources" (Harvard University, 2023) are expected in university level academic writing. The fact that 22 of the 24 test takers in our study made at least one attempt to acknowledge a source indicates the task elicited citation skills resembling those appropriate for the target language use domain. However, our test takers' inconsistent use of quotation marks and author names seems to align with previous research claiming limited language ability and/or lack of training may prevent language learners from properly acknowledging sources (Chan & Yamashita, 2022; Cumming et al., 2006; Gebril & Plakans, 2009; Shi, 2004).

The differences in stylistic conventions elicited by the two forms, namely more generic references and fewer instances of missing citations on Form 2, led us to look more carefully at the reading passages on each form. As a result, we realized that the passages on Form 2 included more mentions of studies/research than those on Form 1. Moreover, those references were often generic, such as, "A 2020 survey found," and did not include the use of quotation marks (which only one of the test takers used). Previous research, such as Gebril and Plakans' (2009) and Plakans and Gebril's (2012) investigations into second language integrated writing tasks, indicates that second language writers sometimes utilize the reading passages in such tasks for language and organizational support. It is possible, therefore, that the way in which sources were cited in the reading passages on the test could have contributed to whether and how test takers chose to acknowledge sources.

Language transformation

Multiple language transformation techniques were observed, even at the B1 level. Even minimal paraphrasing and summaries that incorporated some copied language showed that test takers were attempting to transform the sources to combine them with their own ideas. Although we were not privy to the processes or motivations of the test takers, we viewed these partial transformations as attempts to "observe proper academic conventions," like the patchwriting of students in Howard's (1992) paper.

While previous research indicates that summarizing is a more complex skill than paraphrasing and is more commonly used by higher proficiency writers (Cumming et al., 2006), it is a skill that progresses along a spectrum. In our sample, we observed a nearly equal number of B1 (6) and B2 (5) responses that included summaries on Form 1. However, the proportion of B2 responses that included summaries was higher, and the summaries in the B2 responses revealed the ability to transform language and synthesize the main points (Hedgcock & Ferris, 2009, p. 185) from different parts of a passage if not always between passages.

Overall, it seems the instructions to include information from the sources encourage test takers to attempt to use academic writing subskills, such as quoting, paraphrasing, summarizing, and incorporating ideas from multiple sources. Again, this helps support the authenticity of our task since undergraduate university writers are expected to and often employ multiple language transformation techniques in their writing (Lee et al., 2018; University of Illinois at Urbana-Champaign, 2019–2020).

Rhetorical purpose

In both Lee et al.'s (2018) study on citation practices of L2 first-year university students and Petrić's (2007) study of L2 graduate students' theses, attribution, relabeled as reporting in our study, was the most frequently used rhetorical purpose, with "rhetorically more complex citation types" (Petrić, p. 247) occurring far less frequently. This was not the case in our test takers' writing, where reasons and examples were the most frequently used type. However, as Petrić (2007) points out, attribution may be typical in university writing "since this citation function helps writers display their knowledge of the topic" (p.

247), whereas the primary focus of our task is for test takers to state and support *their* opinion on the given topic, not to merely restate the ideas of others. Moreover, it is important to reiterate that the topics on our exam are suitable for secondary students and thus likely less complex than those included in studies of university level L2 writers. The linguistic complexity of most of our test takers' responses fell into the B1 to Low B2 range, so again less complex than in the papers analyzed in other studies. Nevertheless, it seems clear that the test was eliciting some of the variety in rhetorical purposes desired at higher levels of academic writing, even if the ideas were communicated in a less sophisticated manner.

As mentioned previously, the prompt wording differed on the two forms. Form 1 had an explicit requirement to include information from both source passages, whereas Form 2 did not (Appendices A and B). This may have affected the way test takers engaged with the passages and used them in their responses. Without the requirement to use both passages, there may have been less pressure on test takers to make connections between the two sources, at least partially explaining the reduction in linking and counterargument in the responses on Form 2. This is something to consider in conjunction with the test construct. The CEFR states that B2 language users can "synthesise information and arguments from a number of sources" (Council of Europe, 2020, p. 68), and introductory university writing courses emphasize the ability to utilize multiple sources (New York University (n.d.); University of Illinois at Urbana-Champaign, 2019–2020). In fact, one of the reasons we included two reading passages on our exam was to encourage test takers to display this very ability, among others. We will need to review responses on other forms to determine whether an explicit requirement to refer to both passages is necessary.

Limitations

There were several limitations to this study. Because responses came from an operational exam, we did not collect information on what previous experience test takers had in using and citing sources or what their thought processes were when incorporating the sources into their responses. Those and other variables could have impacted the source

integration patterns we observed in the responses and could theoretically have a different impact on their writing in a classroom setting. Moreover, the test takers were all from a very specific demographic, so the patterns produced in this group cannot be generalized to other testing populations. Another group of test takers, with different training in academic writing, would likely produce different patterns. Finally, we were not always in agreement regarding how something should be coded, particularly for somewhat subjective categories like paraphrase or reason versus example (Table 1). For L2 writers who are still developing their linguistic skills, it can be difficult to quantify their level of success with a particular subskill or infer the reasoning behind their organizational patterns or choice to use an excerpt from a source, particularly when there are no signal words to guide the reader. However, we did not want to limit our analysis by only focusing on features of writing that could be easily quantified or coded.

Conclusion

We investigated the ways in which test takers integrated sources into their writing to determine whether the final item on our academic writing exam was eliciting the type of writing that might be expected in an English-medium university setting. We broke our analysis down by identifying the *stylistic conventions, language transformation,* and *rhetorical purposes* used in test taker responses on two forms of the test. In the two administrations of the exam, they demonstrated the ability to use a wide variety of citation and integration techniques associated with university-level academic writing, even when they were not explicitly instructed to, although they did not always do so with university-level sophistication. These results suggest that even within the context of a timed, standardized writing exam, L2 writers can be encouraged to produce responses that resemble the type of writing expected in an academic setting.

Distinct patterns emerged in the responses on the two forms. While some of these differences could be explained by teaching and learning that took place between administrations, future research could look at how reading passages and instructions might influence the way in which test takers cite and integrate sources into their responses. Findings could be used to help inform the development of reading passages

that can either serve as more accurate models of the readings test takers will encounter in academic settings and/or as models of the writing they will be expected to do, turning the test taking experience into a better learning opportunity.

A summary of findings from this study was passed on to the organization that used the test so that they could address gaps in test taker abilities, such as using quotation marks, successful paraphrasing, and more. In other contexts where similar communication is appropriate, the results of standardized tests can support learning in much the same way that classroom assessments do.

Finally, just as we built upon the frameworks and approaches used by Petrić (2007) and Lee et al. (2018), test administrators or teachers can replicate such studies to systematically check whether writing tasks are eliciting the type of source integration that is suitable for their context. With appropriate scaffolding, L2 writers can even be taught to use these frameworks to analyze their own or their peers' writing.

Author disclosures

The authors are both employed by Second Language Testing Inc., the publisher of the *Written Communication Exam for Academia*. Liubou Shefarevich managed the development and administration of the exam, and Amber Vanderwoude was one of the item writers. The exam was developed by SLTI at the request of a user organization in Japan, with which the authors have no association.

The authors received no specific funding for carrying out the research.

The two authors contributed evenly to the conduct of the research and the writing of the article.

References

American Psychological Association. (2020). *Publication manual of the American Psychological Association 2020: The official guide to APA style* (7th ed.). American Psychological Association. https://doi.org/10.1037/0000165-000

Cambridge (2023). *English vocabulary profile*. Cambridge University Press & Assessment. https://www.englishprofile.org/wordlists

Carr, N. T. (2011). *Designing and analyzing language tests*. Oxford University Press. https://doi.org/10.1017/s0272263112000800

Council of Europe. (2020). *Common European framework of reference for languages: learning, teaching, assessment. Companion volume with new descriptors*. Council of Europe. https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/16809ea0d4

Chan, S., & Yamashita, J. (2022). Integrated writing and its correlates: A meta-analysis. *Assessing Writing,* 54. https://doi.org/10.1016/j.asw.2022.100662

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

Cumming, A., Kantor, R., Baba, K., Eouanzoui, K., Erdosy, U., & James, M. (2006) *Analysis of discourse features and verification of scoring levels for independent and integrated prototype written tasks for the new TOEFL®*. TOEFL Monograph Series MS-30. Educational Testing Service. http://doi.org/10.1002/j.2333-8504.2005.tb01990.x

Gebril, A. & Plakans, L. (2009) Investigating source use, discourse features, and processes in integrated writing tasks. *Spaan Fellow Working Papers in Second or Foreign Language Assessment,* 7, 47–84.

Harvard University. (2023). Harvard College Writing Program, Expos 20. Retrieved March 21, 2023, from https://writingprogram.fas.harvard.edu/expos-20

Hedgcock, J. S., & Ferris, D. R. (2009). *Teaching readers of English: Students, texts, and contexts*. Routledge. https://doi.org/10.4324/9780203880265

Hirvela, A., & Du, Q. (2013). "Why am I paraphrasing?": Undergraduate ESL writers' engagement with source-based academic writing and reading. *Journal of English for Academic Purposes*, *12*(2), 87-98. https://doi.org/10.1016/j.jeap.2012.11.005

Howard, R. M. (1992). A plagiarism pentimento. *Journal of Teaching Writing, 11*(2), 233–245.

Iwata, A. (2012). The effectiveness of summary-writing activities on the improvement of Japanese high school EFL students' writing abilities. *Research Bulletin of English Teaching*, *9*, 1–22.

Kato, M. (2018). Good and poor summary writers' strategies: the case of Japanese high school EFL learners. *Journal of Language Teaching and Research*, *9*(6), 1199–1208. http://doi.org/10.17507/jltr.0906.09

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

Lee, J. J., Hitchcock, C., & Casal, J. E. (2018). Citation practices of L2 university students in first-year writing: Form, function, and stance. *Journal of English for Academic Purposes*, 33, 1–11. https://doi.org/10.1016/j.jeap.2018.01.001

Massachusetts Institute of Technology. (2014). CI-H description & criteria for faculty and instructors/departments and administrators. Retrieved March 10, 2023, from https://registrar.mit.edu/sites/default/files/2019-01/ci-h_criteria.pdf

Moore, Y. (2015). *Investigating valid constructs for writing tasks in EAP tests for use in Japanese university entrance examinations*. British Council Research Reports. https://www.britishcouncil.org/exam/aptis/research/publications/investigating-valid-constructs-writing-eap-tests

New York University. (n.d.). New York University College of Arts and Sciences Expository Writing Program course descriptions, First Year Courses. Retrieved March 21, 2023, from https://cas.nyu.edu/ewp/ewp-courses/course-descriptions.html#3

O'Conner, C. & Joffe, H. (2020). Intercoder reliability in qualitative research: Debates and practical guidelines. *International Journal of Qualitative Methods*, *19*, 1–13. https://doi.org/10.1177/1609406919899220

Petrić, B. (2007). Rhetorical functions of citations in high- and low-rated master's theses. *Journal of English for Academic Purposes*, *6*(3), 238–253. https://doi.org/10.1016/j.jeap.2007.09.002

Plakans, L. & Gebril, A. (2012). A close investigation into source use in second language writing tasks. *Assessing Writing, 17*(1), 18–34. https://doi.org/10.1016/j.asw.2011.09.002

Shi, L. (2004). Textual borrowing in second-language writing. *Written Communication*, *21*(2), 171-200. https://doi.org/10.1177/0741088303262846

Shiratori, K. (2019). Supporting English education reform in Japan: The role of B1 Preliminary. *Cambridge English Assessment – Research Notes,* 73. https://www.cambridgeenglish.org/Images/537871-research-notes-73.pdf

The University of Chicago. (2017). *Chicago manual of style* (17th ed.). The University of Chicago Press. https://doi.org/10.7208/cmos17

University of Illinois at Urbana-Champaign Undergraduate Rhetoric Program. (2019–2020). Rhetoric Instructor Handbook. https://english.illinois.edu/system/files/2020-04/1%20Rhetoric%20Instructor%20Handbook%202019-2020%20.pdf

University of Southern California Dana and David Dornsife College of Letters, Arts and Sciences. (n.d.). WRIT 150 Learning Objectives. Retrieved May 30, 2024, from https://dornsife.usc.edu/the-writing-program/w150-semester-at-a-glance /

Yamada, K. (2003). What prevents ESL/EFL writers from avoiding plagiarism? Analyses of 10 North American college websites. *System, 31*(2), 247–258. https://doi.org/10.1016/S0346-251X(03)00023-X

Appendix A

Form 1 Writing Task with Instructions

You are taking an environmental science course. Your professor asks you to write an online post answering the following question:

In your opinion, does a community benefit more when national parks charge entrance fees or when they provide free entry?

Read the posts written by two of your classmates and then write your own post.

In your post, you **must:**

1. Include information from **both** of your classmates' posts.
2. Clearly state your opinion and support it with reasons.

Write your post in the space below.

You will have 17 minutes to complete the task.

Your post should be at least 100 words.

www.stemcentralcollege.edu/environmentalscience201/discussion_board

Charge Them

by L. Smith Monday, December 4, 9:04 PM

In 2016, Plitvice Lakes National Park in Croatia received nearly 1.5 million visitors, while Yosemite National Park in the United States received around 5 million. These parks, and others throughout the world, have been set aside because of their natural beauty. However, protecting them while supporting the growing number of visitors is expensive. In fact, the US National Park Service currently has an annual budget of about $2.5 billion.

Entrance fees provide essential funding for many parks. They can be used to restore habitats and add safety features, like fences, that protect wildlife. Fees are also used on

infrastructure projects, like the roads and bridges that allow visitors to explore the parks. Visitor centers, staff salaries, and recreational activities are also financed with these fees.

Charging an entrance fee can also reduce crowding, which has a negative environmental impact on many parks. Entrance fees help limit visitors to those who are willing to financially support the parks by paying to enjoy them.

Parks for All

by S. Kim Monday, December 4, 9:53 PM

National parks often represent the best that nature has to offer in a country: the clearest lakes, the tallest waterfalls, the most spectacular mountains. This beauty is a community resource that should be available to all citizens. However, high entrance fees prevent many citizens from enjoying these parks.

A study conducted in the USA showed that wealthy families were three times more likely to visit a national park than low-income families. Taking time off work and paying for transportation to a park—fuel for a car or a bus ticket—may already be difficult for families with limited budgets. Having to pay an entrance fee as well makes a visit even less affordable.

Of course parks require money to operate. However, most of them receive government funding that is paid for with tax dollars. So, many citizens are actually paying for national parks when they pay their taxes. They shouldn't be expected to pay an additional fee in order to visit them.

Appendix B

Sample of Form 2 Instructions

You are taking a digital media course. Your professor asks you to write an essay answering the following question:

Do you think social media algorithms are beneficial or harmful?

Read two articles about social media algorithms and then write your essay.

In your post, you **must:**

- **Include information from the articles.**
- **Clearly state your opinion and support it with your own reasons and examples.**

You will have 20 minutes to complete the task.

Your post should be at least 100 words.