



Castledown

 OPEN ACCESS

# Language Education & Assessment

ISSN 2209-0959

<https://www.castledown.com/journals/lea/>

*Language Education & Assessment*, 6(1), 19–37 (2023)

<https://doi.org/10.29140/lea.v6n1.1175>

## Memorizing and Fabricating? Uncovering High-Stakes Writing Test Preparation



JIEUN KIM

*University of Hawaii at Manoa, USA*

*jieun247@hawaii.edu*

### Abstract

With high-stakes decisions relying on standardized language tests (Chapelle, 2021a), many test-takers seek test preparation from commercial “cram schools,” including TOEFL iBT writing (E.-Y. J. Kim, 2017; S. Kim, 2021). While previous studies have relied on interviews and surveys to investigate Korean students’ test preparation practices, the present study directly observed ten prerecorded online lectures from two popular cram schools with a combined revenue exceeding \$150 million in 2019. The analyzed lectures were publicly available, representing their typical instructional practices. The instructional content of the lectures was coded based on Messick’s (1982) taxonomy of test preparation, revealing three types of instructional activities: Type 1 (enhancing construct-relevant factors), Type 2 (reducing construct-irrelevant factors), and Type 3 (enhancing construct-irrelevant factors) test preparation activities. Type 1 activities primarily covered vocabulary, grammar, and organizational patterns. Type 2 focused on teaching test formats, TOEFL trends, time management, and topic choice. Type 3 addressed memorizing templates, fabricating examples, gaming the scoring system, and providing cramming tips. The lectures seemed to pay enough attention to enhancing construct-irrelevant factors as much as enhancing construct-relevant factors. To increase the test scores, students were encouraged to utilize memorized templates and fabricate examples. However, these practices may undermine support for the explanation and consequence implication inferences of the validity argument of TOEFL iBT. Particularly, this test preparation method raises concerns about whether scores, relying on memorized templates and fabricated evidence, truly reflect academic writing ability. This study provides valuable insights into language testing preparation, applicable to various contexts with different tests.

**Keywords:** test preparation, high-stakes writing test, TOEFL iBT, writing instruction, argument-based validity, second language writing

**Copyright:** © 2023 Kim. This is an open access article distributed under the terms of the Creative Commons Attribution Non-Commercial 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within this paper.

## Introduction

As a certain level of proficiency in the language of instruction is considered to be crucial for academic success, almost all Anglophone universities require evidence of English language proficiency in their international admissions decisions (Chapelle, 2021a). To meet the requirements of English-medium institutions, many prospective international students who are users of English as a second language (L2) take large-scale standardized language tests such as the TOEFL iBT (hereafter TOEFL). The significance of high-stakes language tests leads to intensive language test preparation, which has become a massive enterprise (Burton, 2020; Ross, 2008). A large part of the industry is composed of commercialized preparation programs (Liu, 2014) teaching test-taking strategies and test formats to test takers (Yu et al., 2017). Preparation programs concentrating on test-taking strategies, however, have been questioned regarding both their effectiveness in improving test performance (Winke & Lim, 2017) and their contribution to a gap between test scores and academic performance in college (Zhang-Wu & Brisk, 2021).

Koreans account for a substantial fraction of these test takers and attend commercialized cram schools called *hagwon* to develop test-taking strategies for the TOEFL writing sections (E.-Y. J. Kim, 2017; S. Kim, 2021; Malone & Montee, 2014; Roberts, 2002). In Korean test preparation programs, test takers mainly learn test-taking strategies like memorizing templates rather than developing communicative language skills (E.-Y. J. Kim, 2017; S. Kim, 2021). In addition to face-to-face programs, private cram schools offer prerecorded online courses, known as *ingang*. Online courses are more accessible for students because they can take classes whenever they want, wherever they are. They are potentially more lucrative for *hagwons* since they can attract students, including those who live far from their *hagwons*. As a result, more students are engaged in *hagwon*-based test preparation than are seen entering the school buildings. Language test preparation accounts for a substantial portion of Korea's *ingang* market, which was estimated at \$1.2 billion (Lee, 2018). However, despite the large market for online courses, little research has directly investigated such courses.

This study investigated what is taught in the online TOEFL writing preparation courses at popular cram schools in Korea. In contrast to previous studies that have examined face-to-face courses based on survey and interview data, this study directly observed and analyzed instructional methods of online courses. Consequently, the findings of the study contribute to a deeper understanding of the landscape of language testing preparation practices in English as a Foreign Language (EFL) contexts and how washback from high-stakes writing tests may impact writing instruction.

## Literature Review

### Test Preparation and Performance on High-stakes Language Tests

Test preparation, aimed at improving test scores rather than language skills, can be classified into three types based on Messick's (1982) framework. Type 1 focuses on enhancing construct-relevant aspects of language ability, such as reading fluency. Type 2 aims to reduce construct-irrelevant factors like anxiety, while Type 3 focuses on improving construct-irrelevant test-taking skills, such as memorizing model answers. Types 1 and 2 are considered to contribute to score validity and an accurate representation of the target construct, while Type 3 may inflate scores and pose a threat to validity. Xie (2013) found that Chinese university students engaged more in Type 2 and 3 test preparation, but any type of test preparation did not significantly improve test scores. Similarly, Knoch and colleagues (2020) discovered that test takers generally started with Type 2a test preparation (test familiarization). Most test takers then quickly turned to Type 3 (test-wiseness) whereas a few spent all their time on Type 2 test preparation (e.g., test practice) without engaging in Type 3. Many test takers

never reported engaging in Type 1 test preparation (language learning), or only reported this for one of the subskills (e.g., listening).

In spite of the attention given by many researchers to high-stakes test preparation (Green, 2007; Song & Cheng, 2006), the impact of test preparation on test performance and language instruction remains uncertain, with mixed findings in the literature. Some studies argue that test preparation has a small but significant effect on test performance (Liu, 2014; Xie, 2013), while others find no significant effect (Winke & Lim, 2017). According to Winke and Lim, listening test scores did not differ statistically based on the types of instruction (i.e., practice tests, practice tests, and test preparation instruction). However, the study did not specify the content of the test preparation instruction, and it may not have been sufficiently intensive. The effects of test preparation on receptive skills might also differ from those on productive skills. Regarding production skills tests, such as speaking tests, the use of memorized scripted responses has raised concerns. Raters in a study by Burton (2020), especially those who are familiar with certain test preparation practices, were able to detect rehearsed responses, with disfluency being associated with authenticity. Gates and Cox (2020) highlighted the potential compromise to validity when rehearsed responses are used in oral proficiency interviews. They suggested comparing responses from tasks that could be rehearsed with spontaneous tasks to identify scripted responses. Time-pressured speaking tasks may reveal scripted responses more easily, while written production tasks with more time allowances may be more susceptible to undetected memorized scripts.

### **Test Preparation for TOEFL**

The TOEFL intends to test “what it takes to succeed in an English-speaking academic environment” (Education Testing Service, n.d., para. 1). Due to its high stakes, the test prompts intensive test preparation (Alderson & Hamp-Lyons, 1996; Yu, et al., 2017). Liu (2014) found that half of 14,593 Chinese TOEFL test takers attended cram schools, with test preparation showing a weak relationship with reading and listening scores and a negligible relationship with writing and speaking scores. Test preparation strategies in the writing section commonly include taking practice tests and using templates. As the third largest group of international students enrolled in higher education institutions in the United States (Institute of International Education, 2022), Koreans are among the highest numbers of TOEFL test takers (Choi, 2008; Jeon, 2010; Malone & Montee, 2014; Zhang, 2008). Korean students are also known to engage heavily in test coaching due to their test-oriented culture (Booth, 2018), raising concerns about overreliance on test-taking strategies rather than genuine language ability (Choi, 2008) and doubts regarding discrepancies between writing ability and TOEFL writing scores (Malone & Montee, 2014).

The writing section of the TOEFL aims to assess the ability to “write in English on a wide range of academic and nonacademic topics with confidence and clarity,” according to its performance descriptor (Education Testing Service, 2021, p. 4). However, studies on the test preparation practices of Korean test takers have raised concerns that test preparation practices might result in temporary or otherwise superficial learning. E.-Y. J. Kim (2017) analyzed internet forums where Korean TOEFL test takers exchange information and experiences about the tests and found that the most common strategies used by Korean test takers were templates and TOEFL preparation courses at cram schools. S. Kim (2021) conducted interviews with Korean university students and also observed the prevalence of attending cram schools for test preparation. While some test takers perceived direct test preparation as effective for TOEFL writing tasks, S. Kim (2021) argued that it did not necessarily improve their overall English writing skills, judging from their writing assignments.

Despite these research endeavors, further investigation is still needed to gain a better understanding of the nature of test preparation (Xie, 2013). Firstly, previous studies on test preparation have mostly

relied on indirect methods (e.g., surveys and interviews), with scant research directly observing test preparation courses. Closely inspecting instructional practices can provide richer insights into test preparation, thereby validating previous findings. Secondly, meticulous observation of the test preparation course is necessary as a steppingstone to evaluate the impact of test preparation on test takers' writing test performance and long-term development. To address this gap, the present study examined online TOEFL writing test preparation courses, representing typical writing instructional practices in many language testing contexts. This study was guided by the following research question:

What type of test preparation, and to what extent, do Korean test takers receive through popular online TOEFL iBT writing preparation courses?

### **The Present Study**

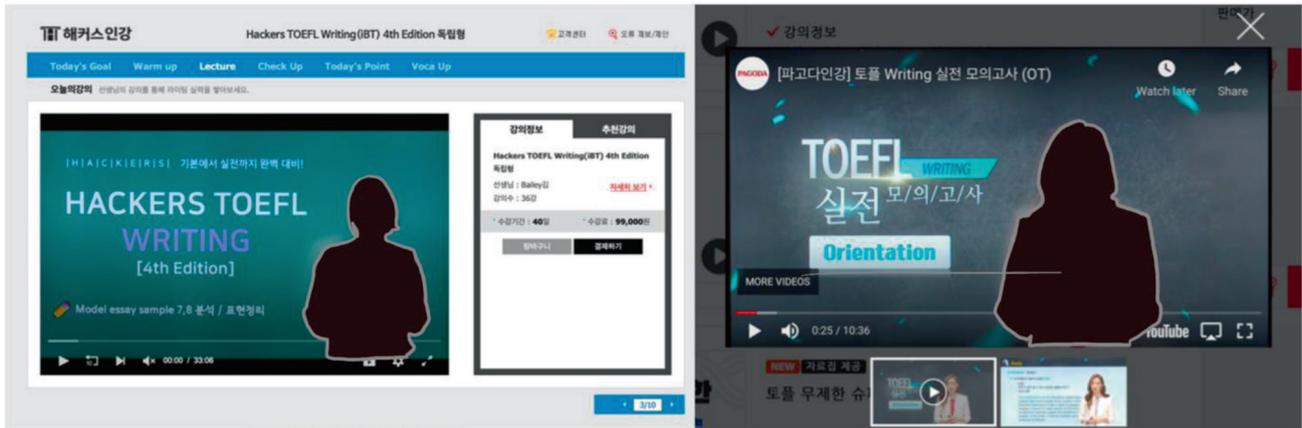
This study examined ten online lectures offered at two giant cram schools in Korea. Located in an expensive neighborhood of Seoul, Hackers Education Group (hereafter Hackers) and Pagoda Academy (hereafter Pagoda) cram schools are the two most lucrative language teaching businesses in Korea. In 2019, Hackers and Pagoda respectively generated more than \$100 million and \$57 million in revenue (Jo, 2020). The accumulated number of Hackers students, exclusively in its language education division, reached three million in 2021 (Hackers, 2021). These schools provide options for online and face-to-face courses. Recently, more learners have attended courses remotely than in-person, and the market for online learning has expanded rapidly since the recent pandemic (Singh, 2022). While face-to-face courses employ instructional methods similar to online courses, the latter, especially those publicly available, serve as prototypes that cram schools present as exemplary lectures. Online lectures are usually taught by designated “star lecturers,” considered the faces of the cram schools. Therefore, examining online courses of these two cram schools (i.e., Hackers and Pagoda) may have implications for face-to-face courses as well. Hackers offered four TOEFL writing preparation courses labeled by proficiency in writing (e.g., Hackers Basic, Hackers Advanced), and Pagoda offered six courses labeled by target score (e.g., Pagoda Hit 70, Pagoda 80 Plus). Each test preparation course consisted of 33.5 lectures on average, which means it takes approximately 22 hours to view all the course lectures (Table 1). Further, between lectures, students are expected to answer the prompts on their own and memorize key vocabulary and templates, potentially spending another 22 hours or more preparing for the test. Although both Hackers and Pagoda had a forum where students could ask instructors questions, there was no interactive activity that provided feedback on students' writing.<sup>1</sup>

In this study, all ten publicly available online lectures, posted to promote both online and in-person courses, were analyzed. These lectures were carefully selected by cram schools to exemplify what students can anticipate when taking their courses. The videos were complete lectures taken from the full course sequences, covering different stages of each course. Therefore, I assumed that the analyzed lectures could represent the typical teaching practices of online test preparation courses. Also, several Pagoda courses included separate orientation videos that discussed the course overview and why test takers might want to take them. Despite its short length, the orientation video provided an overview of instructional practices. For example, in an orientation video, the Pagoda Starter instructor said, “In the writing section, there is a thing called a template...” Since orientations have the potential to provide insight into the teaching practices of the overall courses, they were considered in the analysis. Figure 1 shows example scenes of the lectures.

<sup>1</sup> Many popular instructors hire part-timers to answer students' questions on the forums, so there is little direct interaction between teachers and students even asynchronously.

**Table 1** Number and Length of Lectures in Online Courses

Schools	Number of Lectures		Length of Lectures	
	Mean	Min – Max	Mean	Min – Max
Hackers	42.5	(25 – 52)	43 minutes	(30 – 50) minutes
Pagoda	28.5	(6 – 63)	36 minutes	(17 – 66) minutes
Total	33.5		39 minutes	



**Figure 1** Example Scenes of Lectures from Hackers (Left) and Pagoda (Right)

**Table 2** Characteristics of Lectures Analyzed in This Study

Course	Length	Topics of Lecture
H1 <sup>a</sup>	48m	Explaining a grammar form
H2	41m	Analyzing sample questions and essays
H3	33m	Analyzing sample questions and essays
H4	35m	Explaining how to approach a topic, brainstorm, make an outline, and apply templates
P1	Orien 9m <sup>b</sup> Main 25m	Talking about why the test takers should take the course Explaining how to brainstorm ideas
P2	Orien 13m	Talking about why the test takers should take the course
P3	Main 34m	Analyzing sample essays and applying templates to writing
P4	Orien 10m Main 49m	Talking about why the test takers should take the course Introducing the item types and explaining how to brainstorm and apply templates to writing
P5	Main 29m	Introducing the item types and explaining scoring criteria, useful expressions, how to approach a test item, and how to apply templates to writing
P6	Orien 11m Main 32m	Talking about why the test takers should take the course Explaining how to build an overall essay structure, time management, scoring criteria, and how to apply templates to writing
Total	6h 28m	

<sup>a</sup> H1 = Hackers Basic, H2 = Hackers Intermediate, H3 = Hackers Advanced, H4 = Hackers Actual Test, P1 = Pagoda Starter, P2 = Pagoda Grammar, P3 = Pagoda Hit 70, P4 = Pagoda 80 Plus, P5 = Pagoda Hit 90, P6 = Pagoda Actual Test.

<sup>b</sup> Orien: orientation, Main: main lecture.

I transcribed 6 hours and 28 minutes of videos from ten lectures taught by five instructors. Then, I analyzed the instruction monologue by the unit of an episode, which is a paragraph-sized semantic unit in oral discourse (Johnstone, 2017; Van Dijk, 1981). Each episode was often signaled by pauses, hesitations, and markers indicating a shift in topic or perspective. For example, a shift in the topic was signaled by markers such as *어, 그런데* (um, in fact). Sometimes, the instructors' use of a student's voice like "*선생님, 저 아는 게 없어요* (Teacher, I know nothing)" showed the changes in perspective and initiated another episode. First, I reviewed the videos twice and identified fifteen emerging codes. Then, I eliminated codes that were less relevant to test preparation itself, such as advertisements, while others were either combined or revised. For instance, codes related to grammar instruction, such as "grammar explanation" and "teaching typical grammatical errors," were merged into a single code. The videos were then subjected to three additional rounds of review, iteratively revising the codes. This process yielded the twelve codes categorized into three types of test preparation based on Messick's (1982) types of test preparation: Type 1 (enhancing construct-relevant factors), Type 2 (reducing construct-irrelevant factors), and Type 3 (enhancing construct-irrelevant factors) test preparation. Table 2 shows the characteristics of the lectures analyzed in this study.

## Findings

The test preparation episodes identified from the observed lectures are summarized in Table 3. For Type 1 test preparation, vocabulary, grammar, and different organizational patterns were taught. Regarding Type 2 test preparation, the instructors explained test formats, TOEFL trends, time management, and topic choice. Introducing templates, how to make up examples, techniques to game the scoring criteria, and cramming tips were Type 3 test preparation commonly employed in the courses.

**Table 3** *Test Preparation Episodes Provided in Lectures*

Test Preparation Episodes	H1 <sup>a</sup>	H2	H3	H4	P1	P2	P3	P4	P5	P6	Total (%)
<b>Type 1</b>											
Vocabulary	7	30	35	3			7	5	2	5	94 (35)
Grammar	2	2	2	1			4	6	1	2	20 (7)
Organizational pattern				1	1		1	1			4 (1)
Subtotal	9	32	37	5	1		12	12	3	7	118 (44)
<b>Type 2</b>											
Test formats					1				2	1	4 (1)
TOEFL trends	1		1	1				1	3		7 (3)
Time management				1	3			7	1	2	14 (5)
Topic choice		1	1		3			3	6		14 (5)
Subtotal	1	1	2	2	7			11	12	3	39 (15)
<b>Type 3</b>											
Templates		11	1	5	6		7	10	7	8	55 (21)
Fabricating examples		1		5	1			1	3	1	12 (4)
Gaming the scoring system	1	3		6		3	1	9	1	5	29 (11)
Cramming tips	1	4	2				1	3	2		15 (6)
Subtotal	2	19	3	16	7	3	9	23	13	14	111 (41)
Total	12	52	42	23	15	3	21	46	28	24	268

<sup>a</sup> H1 = Hackers Basic, H2 = Hackers Intermediate, H3 = Hackers Advanced, H4 = Hackers Actual Test, P1 = Pagoda Starter, P2 = Pagoda Grammar, P3 = Pagoda Hit 70, P4 = Pagoda 80 Plus, P5 = Pagoda Hit 90, P6 = Pagoda Actual Test.

## Type 1 Test Preparation

Type 1 test preparation aims to enhance linguistic abilities relevant to the test construct. This was the most common type of test preparation, accounting for just under half (44.02%) of all identified teaching episodes. For Type 1 test preparation, the instructors primarily covered vocabulary, grammar, and organizational patterns necessary to increase writing test scores.

### *Vocabulary*

Vocabulary was taught in most lectures, accounting for almost 80% of the Type 1 test preparation episodes. During the lectures, the meaning, usage, and common mistakes of certain vocabulary items were explained. For example, the Hackers Advanced instructor discussed a mistake commonly made by Korean learners of English. Because the word 듣다 (*listen*) collocates with the word 수업 (*class*) in Korean, Korean learners often write “listen to a class” instead of “take a class.” Additionally, the instructors introduced synonyms and encouraged students to use more advanced vocabulary. The Pagoda 80 Plus instructor criticized the phrase “I think” as childish and offered “as far as I am concerned” as a more sophisticated alternative. They recommended using “as far as I am concerned” because it sounds fancy.

### *Grammar*

A few lectures primarily taught grammar forms. Pagoda Grammar was solely devoted to grammar lessons, and the lecture from Hackers Basic focused on relative clauses. In other courses, grammar forms were not the main focus; however, grammar forms were taught incidentally in order to explain model essays and templates. Specifically, the instructors often taught articles (five times) and prepositions (three times), which have been acknowledged to be particularly challenging for Korean EFL learners (Zheng & Park, 2013).

### *Organizational patterns*

The lecturers spent much time explaining different organizational patterns between English texts and Korean texts, as the logic expressed through written text can differ from culture to culture (Kaplan, 2005). Despite criticisms of Kaplan’s contrastive rhetoric in academia (Ferris & Hedgcock, 2014), the instructors tended to claim that the Korean language takes an indirect approach to a topic whereas the English language speaks directly to it. The instructors emphasized that the main idea should be introduced at the beginning of the English essay.

(Excerpt 1: Pagoda Hit 70) 좀 우리나라 사람들은 듣다 듣다 듣다 보면은 결론을 얘기하는 경우가 되게 많은데 미국 애들은 그렇게 참을성이나 인내심이 강하지 않습니다. 여러분 하고 싶은 얘기를 먼저 던지셔야 돼요. You know people in Korea give a conclusion after we listen to a story for a while. These Americans do not have that strong of tolerance or patience. You should start with what you want to say.

In Excerpt 1, the lecturer attempted to explain the reasons for different organizational patterns but ended up perpetuating certain stereotypes. She found reasons for the direct approach to English writing outside of writing. It was attributed to an overly simplified and biased view of Americans’ cultural characteristics.

## Type 2 Test Preparation

The purpose of Type 2 test preparation is to reduce construct-irrelevant interference in test-taking to improve scores. This type of test preparation accounted for 14.55% of the overall test preparation episodes. Training for familiarization with tests can be considered Type 2 test preparation, as demonstrated in Knoch et al (2020).

### *Test formats*

In the lectures, the task types, prompt types, and time limits for each task were introduced to increase familiarity with the test. Test familiarization was usually done in the orientations because it was considered to be a prerequisite for other test preparation processes. Similar information was also repeated throughout the other lectures. For instance, the Pagoda Starter instructor explained that writing tasks are preceded by reading, listening, and speaking sections and discussed what independent tasks and integrated tasks are.

### *TOEFL trends*

Sometimes, the instructors talked about the “recent trend” in prompt types or topics. They mentioned that certain prompt types or topics are popular.

(Excerpt 2: Hackers Actual Test) 요새 이런 주제들이 꽤 많이 나오고 있는데요. 오늘 주제, 사실 점수가 꽤 많이 갈려요. 구체화가 꽤 중요한 주제이고요. 영화, 책에 기반한 영화. What, 이게 나오지 않으면 점수가 고득점이 힘든 주제이기도 했어요. 문제 한번 먼저 살펴보도록 할게요. These topics appear quite often on tests, recently. Today’s topic discriminates between high and low scores pretty well. The specification is critical for this topic. (The topic is) movie, a movie based on a book. Without specifying “what,” it is difficult to get a high score on this topic. Let’s take a look at the question.

According to Excerpt 2, the prompt about a movie based on a book is trending. The Pagoda 80 Plus instructor said “choose-and-defend-a-position-on-an-issue” accounts for 80% of the independent task prompts. Additionally, the Pagoda Actual Test said that choosing a preference out of three options is a prompt type that has been around since 2012, and Hackers Hit 90 also noted that it is a less frequent but challenging prompt. It is unclear what evidence these claims are based on or how the cram schools acquire such knowledge, though it may be worth noting that Hackers was previously prosecuted for collecting and sharing TOEIC test items (Park, 2012). Having instructors and paid informants take the test multiple times and harvesting test prompts, the cram schools can provide a narrower type of test preparation that is tuned to a specific test. Test takers who have been trained by a narrow curriculum can potentially take the test repeatedly, hoping to encounter prompt types or topics they have extensively prepared for, to maximize their “My Best Score.”

### *Time management*

The TOEFL has a 20-minute time limit for integrated writing tasks and a 30-minute time limit for independent writing tasks. In order to use the given time as efficiently as possible, minute-by-minute time management was taught during the test preparation lectures. The instructors agreed that the introduction and conclusion should be written within three minutes so that test takers can invest the rest of the time in writing body paragraphs. In Pagoda 80 Plus, the instructor advised test takers to write an introduction and a conclusion in three to four minutes. The Pagoda Actual Test instructor said that the

introduction should be written in three minutes and the conclusion in one minute. The instructors said writing an introduction and conclusion in that short amount of time is possible if test takers jot down memorized templates.

### *Topic choice*

TOEFL independent writing tasks often include choosing and defending a position on an issue. In response to this prompt, learners might linger on choosing their positions, so instructors explained how to quickly decide whether they agree or disagree with it. They recommended choosing a stance that is easier to approach. Although the explanation might sound vague, they said a position that would change the existing situation is easier to start.

(Excerpt 3: Pagoda Starter) 그래서 여러분들 이 문제를 딱 봤을 때 여러분들이 agree 가 쉬울까요. disagree 가 쉬울까 어떤 주장이 조금 더 글을 쓰기가 쉬울지 한번 잠시만 생각을 해보세요. 주장을 여러분들이 정하실 때는요. 아무래도 변화가 있는 주장이 조금 더 글을 쓰기가 쉽습니다. 변화가 있는 주장이라 그러면은...애를 듣는 게 변화가 있을까요. 안 듣는 게 변화가 있을까요. 여러분들 생각을 해보세요. 당연히 들어야지 변화가 있겠죠. So, when you first look at this question, take a moment to think about whether agreeing is easier or disagreeing is easier to write an essay. When you take a position, the argument that holds changes is easier to write. The argument with changes means...Which causes a change, taking this or not taking this (taking the course or not taking the course)? Everyone, think about it. Of course, there is a change when you take it (the course).

For example, in Excerpt 3, when responding to a prompt asking whether students should take a certain course or not, the instructor advised students to choose the side of “taking the course” because it makes changes to the given situation (i.e., not taking the course represents the default or status quo) and offers more to discuss. According to the instructors, the writing task does not intend to assess the test takers’ knowledge or beliefs, but their linguistic ability. In other words, a response to a writing task need not reflect what a test taker actually believes.

### **Type 3 Test Preparation**

Type 3 test preparation aims to improve test scores by enhancing construct-irrelevant test-taking skills. Memorizing templates, fabricating examples, taking advantage of the scoring system, and using cramming tips were identified as Type 3 test preparation in this study. This type of test preparation accounted for 41.42% of the test preparation episodes, but it could raise the issues of inflating the test scores and posing a threat to test validity (Messick, 1982).

### *Templates*

In line with previous studies (E.-Y. J. Kim, 2017; S. Kim 2021), templates were often taught in test preparation courses. The most common Type 3 test preparation episode is learning templates, accounting for half of all Type 3 test preparation episodes. Templates were taught in almost all of the lectures (Table 3), except for Hackers Basic and Pagoda Grammar, and they were also the main topic of four lectures (Table 2). The templates included formulaic chunks commonly found in the introduction and conclusion paragraphs as well as transition words (see Appendix for two example templates). The instructors proudly presented the templates that they had created and emphasized that students should memorize them. They even criticized other templates for using basic vocabulary and being widely used.

(Excerpt 4: Pagoda Actual Test) 그래서 한국의 중학생부터 해서 한국에 나이 많으신 분들까지 그런 다음에 전 세계적으로 이 template 을 쓰고 있습니다. 여러분들 아무래도 나만의 것 나만의 template 을 만드는 게 중요하겠죠. 그래서 요거는 쓰지 마시고, 이거 쓰고 계셨으니 이거 쓰지 마시고요. 새로운 template 하나 외워봐 주시면 좋을 것 같아요. So, from middle school students in Korea to older Koreans, and then people all around the world are using this template. Everyone, anyhow it is important to make my own, my own template. So, do not use this. You have been using this, so do not use this. Now let's memorize a new template.

In Excerpt 4, the instructor criticized the first template provided in the Appendix and suggested the second template as an alternative. The former template was to be avoided because it was widely used, making it more likely that raters could detect memorized templates. In addition to memorizing key phrases, another type of template presents a frame of logical flow depending on the topic. This type of template was intended to help test takers brainstorm their ideas quickly. These templates were literally called “Cheating Keys” in Pagoda Starter and Pagoda 80 Plus.

(Excerpt 5: Pagoda Starter) 제가 아까 여러분들한테 Cheating Keys 그러니까 좀 기억을 해 주셨다가 시험장에서 문제 딱 보고 이거 되지 않을까라고 생각되는 거 6가지 한번 드려보도록 할게요. 애는 6가지 핵심 논리라고 얘기를 했어요. 그죠, 일단은 첫 번째 우리 문제에서 조금 재미있는 놀이 활동이나 오락거리가 나왔을 때. 그럴 때는 여러분들이 스트레스라고 조금 생각을 해보시면 좋을 것 같아요. 우리 오락거리가 있다. 이러면 어때요. 그 주제를 함으로써 행복해져요. 스트레스가 풀려요. 웃어요. 이런 얘기 하실 수가 있겠죠? 그래서 여러분들이 어? 이거는 뭔가 스트레스 해소할 수 있는 활동이다. 이렇게 생각을 해 보시면 될 것 같고요. I told you about “Cheating Keys” earlier. So, I will present six keys that you can memorize and apply right after you read the prompt at a test center. These are six core logical frames. Right, first if there is a playful activity or entertainment in the prompt, you can think of “stress.” We have entertainment. How does that sound? By doing that topic, we become happier and less stressed. We will laugh more. You can talk about this? So, if you think that this can be an activity that relieves stress. You can approach this way.

In Excerpt 5, the instructor gave an example of “Cheating Keys.” That is, for any entertainment-related prompt, students can write about how the activity can release stress. Using this strategy, test takers can approach different prompts in a certain way on a broader level and then use templates with key phrases to fill in the essay.

### ***Fabricating examples***

TOEFL writing prompts require a test taker to present examples or personal experiences to support one's idea. In case test takers cannot come up with relevant examples, the instructors advised students to fabricate examples, teaching how to make up experiences and even statistics. In Pagoda Hit 90, for instance, the instructor talked about how students can lie about their travel experiences in Thailand even if they have never been there. In Hackers Actual Test, the instructor literally said, “you can make it up” and taught how to make up an example of a movie based on a book. The instructors claimed that the raters evaluate test takers' writing ability and do not care whether the examples are true or not.

(Excerpt 6: Pagoda Actual Test) 그래서 우리가 아무래도 지어내는 거기 때문에 좀 credibility 신용도를 높여줘야 되겠죠. 그러다 보니까 어떤 연구 기관인지 어떤 연구를 진행했는지 이런 부분을 조금만 상세하게 설명을 해 준다고 생각을

하시고요. [...] 그래서 샘플로 우리 한 번 같이 써보도록 할게요. 보니까 첫 번째 문장, 연구기관과 어떤 연구인지 설명을 해야 된다고 그랬어요 그렇죠. 그리고 연구 기관을 우리는 저명하다라고 표현을 할 거예요. 그래서 제가 이렇게 과란색으로 표현을 해놨는데 여러분들 이거를 외워놔 주시고 그런 다음에 검은색 부분을 여러분들이 항상 바꿔주시면 돼요. 그렇죠. 보도록 할게요. [...] According to research conducted by DS, conducted 하면 무슨 뜻이에요? 수행된, 그죠. 그런 다음에 이 기관을 조금 더 신뢰성을 쌓기 위해서 설명이 들어갑니다. “One of the most well-known research centers in South Korea” 이렇게 들어가는데 [...] So, since we are making this up, we need to raise some credibility. Therefore, please think that we elaborate on what kind of research institution conducted a study and what kind of research was in detail. In the first sentence, we explain how a research institution conducted a study. [...] So, we will work on a sample. Let’s see. In the first sentence, we should explain a research institution and what the research is about, right? And we will say that the research institution is prominent. So, I colored this in blue. Please memorize (the blue part). And then, you can just change the black part, right. Let’s see. [...] “According to research conducted by DS.” What does “conducted” mean? Yes, “performed.” Then, let’s add some explanation to build the credibility of this institution. We can put “one of the most well-known research centers in South Korea.”

In Excerpt 6, the instructor introduced fabricating strategies, creating a fictional research institute. She accentuated the status of this fake institute by calling it “prominent” and “one of the most well-known research centers in South Korea.”

### *Gaming the scoring system*

In the lectures, the instructors explained how to make the most of the knowledge of the scoring system. They discussed factors that affect test scores and those that do not, saying that the introduced criteria are based on the official guide (e.g., Education Testing Service, 2020). Some lectures also explained how e-rating and human raters are involved in the scoring. According to the lectures, test takers should refrain from going “off-topic”, because doing so could result in a reduced score. The Hackers Actual Test instructors said it wasn’t the case before, but now raters boldly deduct scores if an essay deviates from the topic. These episodes could arguably be classified as Type 2 test preparation; however, they were extended to gaming the scoring criteria. First, to avoid writing things that are irrelevant to the topic, students are advised not to write three body paragraphs. The Pagoda Actual Test instructor cautioned students not to write more than two body paragraphs because students may write things that are irrelevant to the topic. Hackers Actual Test instructor even suggested writing one detailed and lengthy body paragraph rather than writing two superficial body paragraphs. Moreover, instructors seemed to agree that getting a high score does not depend so much on the conclusion as on the introduction. The Pagoda 80 Plus instructor suggested that writing without an introduction could be considered “off-topic” but a conclusion rarely affects scores. This aligns with advice for allocating more time to writing an introduction than writing a conclusion.

(Excerpt 7: Pagoda 80 Plus) 서론은 너무 중요하고 그런 다음에 여러분들 여기서 (결론에서) 점수가 더 잘 쓴다고 더 길게 쓴다고 올라가지는 않아요. 그렇기 때문에 빨리 끝내주셔야 되거든요. 그런데 서론에서 여러분들이 막 오다가 나 있고 이런 것들 다. 빠져가지고 문법적인 오류를 냈어요. 채점자가 사람인데 그거 보면서 무슨 생각하겠어요? ‘이 사람 영어가 생각보다는 좀’ 이렇게 될 수가 있겠죠. 그런 생각을 다. 차단을 해줘야 되는 게 서론이에요. 그래서 완벽하게 쓸 수 있도록 여러분들이 계속, 계속 써보시면 되겠습니다.

An introduction is vital, and writing a good and long (conclusion) will not increase your score. That is why you should finish it quickly. But if you have typos in the introduction. And many things, all of them are missing. And you made grammatical errors. Raters are humans, too. What would they think? “This person’s English isn’t quite what I expected.” This can happen. Those thoughts should be blocked in the introduction. In order to write a perfect introduction, you have to keep, keep writing.

According to Excerpt 7, gaining a high score requires a strong introduction, while the conclusion does not matter. The Pagoda Actual Test instructor also downplayed the importance of the conclusion and said that test takers could skip it. Similarly, a test taker in E.-Y. J. Kim (2017) reported that she always received high scores without writing conclusions. The TOEFL rubrics (2022), however, do not mention introductions and conclusions but do suggest advanced essays should demonstrate unity, progression, and coherence.

### ***Cramming tips***

Tips to prepare for the test were often discussed due to the characteristics of online courses. Since the classes were taught asynchronously, it is almost impossible for the instructors to interact with students. To motivate students, the instructors introduced some tips for self-study, encouraging students to review the class and study actively. The tips for cramming were closely related to memorizing templates and sample essays, rather than more construct-relevant learning activities like studying vocabulary.

## **Discussion**

The present study set out to investigate the online TOEFL preparation courses in Korea. In the analyzed lectures, Type 1 test preparation accounted for 44%, Type 2 for 15%, and Type 3 for 41%. While Type 1 was the most common type of test preparation, Type 2 and Type 3 test preparations combined took up a greater proportion of the test preparation than Type 1 test preparation alone, as demonstrated in Xie (2013) and Knoch et al. (2020). In addition, as Knoch et al. noted, the types of test preparation episodes were difficult to discern because they were interconnected. Vocabulary and grammar (Type 1), for instance, were primarily taught as tools for learning sample essays and templates (Type 3). Memorizing those templates was more likely to allow students to manage their time efficiently (Type 2) by composing an introduction and a conclusion within a short period of time. Also, explaining scoring criteria that might seem to be Type 2 test preparation was extended to teaching how to game them (Type 3). Choosing a position that could go against the writer’s belief (Type 2) can be closely linked to fabricating examples (Type 3). Each type of test preparation yielded the following implications.

With respect to Type 1 test preparation, knowledge about vocabulary, grammar, and organizational patterns was taught to improve academic writing abilities. There is no doubt that students’ writing performance could be enhanced by a varied vocabulary (Ferris, 1994; Grabe & Kaplan, 1996; Jarvis, et al., 2003) as well as accurate use of grammar (Schoonen et al., 2003; Trapman et al., 2018). As well, language teachers are encouraged to teach vocabulary and grammar knowledge that is appropriate for academic discourse (Coxhead & Byrd, 2007). However, the long-term effects of teaching vocabulary and grammar limited to sample essays and templates need further investigation. In the observed lectures, instructors primarily focused on the vocabulary and grammar necessary for interpreting model essays, neglecting broader language skills needed for producing high-quality academic writing. Additionally, there was a lack of interaction, drafting, feedback, and conferences in the instructional approach. Consequently, the narrowed curriculum based on sample responses and templates may have short-term effects limited to specific test prompts. The effectiveness of this training might not extend to

non-testing contexts, and the impact of restricting vocabulary and grammar knowledge to test-taking situations on a writer's overall language development remains unclear.

Regarding Type 2 test preparation, familiarizing oneself with test formats was considered a precondition for test preparation. Since understanding the types of writing tasks (i.e., independent and integrated tasks), prompt examples, and time limitations was considered a necessary first step to preparing for the high-stakes test, familiarization was included in orientation lectures and repeated throughout the courses. Previous studies confirm the importance of test familiarization in test preparation (Ma & Cheng, 2015; Yu et al., 2017). Familiarizing with test formats has been found to be effective in improving test performance (Winke & Lim, 2017; Zhang, 2008) and reducing the impact of irrelevant factors on test results, leading to more accurate assessments of language abilities (Messick, 1982). Test takers can achieve familiarization through official guides, practice tests provided by test providers, and private test preparation programs.

Type 3 test preparation was largely concerned with teaching templates, making up examples, and gaming the scoring system. Writing with templates was one of the most common test preparation activities, as reported in previous studies (E.-Y. J. Kim, 2017; Liu, 2014; S. Kim, 2021). In the analyzed lecturers, the instructors strived to create and advertise their own templates, competing against one another.<sup>2</sup> Similarly, test takers in E.-Y. J. Kim (2017) frequently discussed some templates available only from specific instructors. Furthermore, instructors encouraged students to fabricate personal experiences and factual information as supporting evidence, emphasizing that the test evaluates writing ability rather than knowledge or truthfulness. Notably, the importance of the conclusion was overlooked, potentially misleading students to believe that academic writing does not value conclusions as much.

This sort of Type 3 test preparation casts doubts on whether test scores based on memorized templates and fake supporting evidence can be attributed to the intended construct, academic writing ability. Test scores should reflect the theoretical components of language proficiency in order to satisfy the explanation inference for the validity of the test (Chapelle, 2008)<sup>3</sup>, especially for TOEFL writing, which aims to assess effective writing skills in English across various academic topics (Education Testing Service, 2021). Some studies provided backing for the explanation and extrapolation inferences of the TOEFL writing. The TOEFL writing tasks were associated with non-test writing samples in terms of textual features (Riazi, 2016) and rhetorical profiles (Beigman Klebanov et al., 2019). Llosa and Malone (2019) also found that students' TOEFL scores were closely related to the first drafts of their writing course assignments and concluded that the TOEFL writing score could represent students' preparation for university writing courses. Those studies, however, did not examine the TOEFL preparation process so they might have overlooked some cases where test takers rely on memorized templates and forged examples to complete the writing tasks. Writing responses based on Type 3 test preparation, if undetected, may not accurately represent the theoretical constructs of writing proficiency or academic writing ability. Relying on templates and fabricated examples may lead to high test scores but may not reflect the genuine academic writing skills required in higher education.

Furthermore, the test preparation observed in this study rebuts the consequence implication inference of the validity of the TOEFL that the test brings a positive impact on society (Chapelle, 2021b). As a social benefit, the TOEFL is intended to prompt "the creation and the use of test preparation materi-

<sup>2</sup> The orientations gave a glimpse of how fiercely instructors compete against one another. In orientations, the instructors stressed their confidence in increasing TOEFL scores in a short period of time. For instance, the Pagoda Hit 90 instructor introduced herself as 만점 제조기 (a perfect score maker).

<sup>3</sup> Chapelle (2021b) recently argued that an explanation inference could be combined with an extrapolation inference for interactionist constructs.

als and activities that would more closely resemble communicatively oriented pedagogy in academic English courses” (Educational Testing Service, 2011, p. 15). However, the test preparation episodes centered on teaching the templates and fictional examples seem far outside of communicative pedagogy and pose the risk of promoting plagiarism. The misunderstanding of plagiarism among English learners is often attributed to their cultural backgrounds (Bloch, 2001; Gu & Brooks, 2008), but they might also be misguided by the test preparation course they took before coming to the United States. These limitations could explain why some international students reported that memorizing templates led to success on the TOEFL but did not enhance communication and writing skills in college, saying “I must have taken a fake TOEFL” (Zhang-Wu & Brisk, 2021, p. 7). S. Kim (2021) also noted that test preparation training from cram schools did not necessarily improve writing performance in class. In addition, if attending test preparation courses is effective or at least perceived as effective in improving test scores, the test might not be fair for all test takers (as some test takers cannot afford preparation courses), or beneficial to the immediate community and broader society.

Lastly, the selected lectures as well as the entire curriculum mainly taught independent tasks rather than integrated tasks.<sup>4</sup> During Pagoda Starter, the instructor said that independent tasks were mainly taught throughout the course because they are “basic, but more difficult.” In fact, compared to independent tasks, whose prompts are highly predictable (E.-Y. J. Kim, 2017), the scores of integrated tasks fluctuate more and are affected by the characteristics of the prompt (Cho et al., 2013). Integrated tasks might be more comparable to academic writing tasks in the target domain in some ways (Cumming et al., 2005; Llosa & Malone, 2019; Taylor & Angelis, 2008); however, they differ from independent tasks in terms of both the product and process of writing (Guo et al., 2013; Knoch et al., 2014; Kyle, 2020; Kyle & Crossley, 2016; Plakans & Gebriel, 2013). Independent tasks assess effective topic addressing, while integrated tasks evaluate the ability to extract and relate information from audio and reading materials (Education Testing Service, 2022). Type 3 test preparation, with templates and fake examples, may be less suitable for integrated tasks compared to independent tasks. Consequently, prioritizing independent tasks may result in higher scores with the same time and effort invested compared to integrated tasks. Notably, as of July 26, 2023, TOEFL no longer includes independent writing tasks (Education Testing Service, n.d.). This change also highlights the limitation of an independent writing task, which is more susceptible to construct-irrelevant test preparation. The findings related to independent tasks remain relevant to test preparation practices for other tests that include an independent writing task.

The findings of the present study serve as a foundational exploration into understanding the impacts of test preparation practices on learners’ test performance and long-term language development. While growth based on Type 1 test preparation might take some time to manifest, it would be intriguing to explore the impact of Type 3 test preparation after completing one or two full courses. Although Winke and Lim (2017) did not find significant effects of test preparation on test takers’ listening performance, their instructional simulation of commercial test preparation was relatively brief, lasting only a total of four hours of training. Given that a single course, as observed in our study, spans an average of 22 hours, the brevity of the training may explain the lack of observable effects. Furthermore, it will be valuable to investigate the effects of test preparation on actual test takers who genuinely need test scores and actively engage in reviewing lectures and memorizing templates, rather than individuals who take the test solely for research purposes. These considerations lay the groundwork for future research, emphasizing the need for more comprehensive and extended investigations into the diverse dimensions of test preparation and its influence on test performance.

<sup>4</sup> Basic courses mainly taught independent tasks rather than integrated tasks. Three Pagoda courses that covered both tasks (i.e., Pagoda Hit 70, Pagoda 80 Plus, and Pagoda Actual Test) spent more time on independent tasks (mean = 414 minutes) than on integrated tasks (mean = 313 minutes). While two Hackers courses (i.e., Hackers Intermediate and Hackers Advanced) had separate sections for independent and integrated tasks, they did not provide the length of each lecture but showed the average length of lecture including both tasks.

## Limitations

While the present study provided valuable insights into an underexplored form of test preparation, it had certain limitations. Firstly, although Hackers and Pagoda are significant cram schools in Korea, only a limited number of courses were analyzed. The study also observed only one lecture from each test preparation course instead of examining the entire courses, which might have offered a more comprehensive understanding of test preparation practices. Additionally, although Hackers reports an accumulated enrollment of three million students, the specific distribution across individual courses remains undisclosed. Gaining insight into the enrollment distribution could better help us comprehend the impact of these test preparation practices. Despite these limitations, the present study contributes to the ongoing discussions surrounding the validity of high-stakes language tests, thereby prompting future research on test preparation and its impact.

## Conclusion

By observing online test preparation courses that may have a remarkable impact on a wide variety of stakeholders, some insights into test preparation for large-scale, high-stakes language tests were gained. In order to increase test scores, commercial cram schools often appear to focus on construct-irrelevant factors rather than construct-relevant factors, using every measure possible. These measures were consumed by test takers who are likely to take several levels of courses for four subskills—as a single course takes 22 hours on average, it might take them 176 hours to complete two courses for each subskill (22 hours \* 2 courses \* 4 subskills). Considering the time, money, and effort invested in test preparation, its impact on the test-taking community and society is not negligible. Furthermore, how Hackers “hacks” the test and how Pagoda teaches “cheating keys” are unlikely to be limited to Korea or the TOEFL iBT alone. Stakeholders of high-stakes standardized tests are therefore urged to consider how test preparation activities bear on support for the explanation and consequence implication inferences in their tests’ validity arguments.

## Disclosure Statement

Author has no competing interests to report.

## References

- Alderson, J. C., & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. *Language Testing*, 13(3), 280–297. <https://doi.org/10.1177/026553229601300304>
- Beigman Klebanov, B., Ramineni, C., Kaufer, D., Yeoh, P., & Ishizaki, S. (2019). Advancing the validity argument for standardized writing tests using quantitative rhetorical analysis. *Language Testing*, 36(1), 125–144. <https://doi.org/10.1177/0265532217740752>
- Bloch, J. (2001). Plagiarism and the ESL student: From printed to electronic texts. In D. Belcher & A. Hirvela (Eds.), *Linking literacies: Perspectives on L2 reading-writing connections* (pp. 209–228). University of Michigan Press.
- Booth, D. K. (2018). Introduction. In D. K. Booth (Ed.), *The sociocultural activity of high stakes standardised language testing: TOEIC washback in a South Korean context* (Vol. 12) (pp. 1–6). Springer.
- Burton, J. D. (2020). “How scripted is this going to be?”: Raters’ views of authenticity in speaking-performance tests. *Language Assessment Quarterly*, 17(3), 244–261. <https://doi.org/10.1080/15434303.2020.1754829>

- Chapelle, C. A. (2021a). Looking ahead to the next generation of academic English assessments. In X. Xiaoming & J. M. Norris (Eds.), *Assessing academic English for higher education admissions* (pp. 200–214). Routledge.
- Chapelle, C. A. (2021b). Uses and consequences of the test scores. In C. A. Chapelle (Ed.), *Argument-based validation in testing and assessment* (pp. 38–53). Sage Publications.
- Cho, Y., Rijmen, F., & Novák, J. (2013). Investigating the effects of prompt characteristics on the comparability of TOEFL iBT™ integrated writing tasks. *Language Testing*, 30(4), 513–534. <https://doi.org/10.1177/0265532213478796>
- Choi, I. C. (2008). The impact of EFL testing on EFL education in Korea. *Language Testing*, 25(1), 39–62. <https://doi.org/10.1177/0265532207083744>
- Cohen, A. (2006). The coming of age of research on test-taking strategies. *Language Assessment Quarterly*, 3(4), 307–331. <https://doi.org/10.1080/15434300701333129>
- Coxhead, A., & Byrd, P. (2007). Preparing writing teachers to teach the vocabulary and grammar of academic prose. *Journal of Second Language Writing*, 16(3), 129–147. <https://doi.org/10.1016/j.jslw.2007.07.002>
- Cumming, A., Kantor, R., Baba, K., Erdosy, U., Eouanzoui, K., & James, M. (2005). Differences in written discourse in independent and integrated prototype tasks for next generation TOEFL. *Assessing Writing*, 10(1), 5–43. <https://doi.org/10.1016/j.asw.2005.02.001>
- Educational Testing Service. (2011). *Validity evidence supporting the interpretation and use of TOEFL iBT scores* (Vol. 4). ETS. [https://www.ets.org/s/toefl/pdf/toefl\\_ibt\\_insight\\_s1v4.pdf](https://www.ets.org/s/toefl/pdf/toefl_ibt_insight_s1v4.pdf)
- Education Testing Service. (2020). *Performance descriptors for the TOEFL iBT® test*. <https://www.ets.org/content/dam/ets-org/pdfs/toefl/toefl-ibt-performance-descriptors.pdf>
- Education Testing Service. (2021). *Official guide to the TOEFL iBT test* (6th ed.). McGraw-Hill Education.
- Educational Testing Service. (2022). *Scoring guides (rubrics) for writing responses*. ETS. [https://www.ets.org/s/toefl/pdf/toefl\\_writing\\_rubrics.pdf](https://www.ets.org/s/toefl/pdf/toefl_writing_rubrics.pdf)
- Education Testing Service. (n.d.). TOEFL iBT: About the Test. ETS. <https://www.ets.org/toefl/score-users/ibt/about.html>
- Ferris, D. R. (1994). Lexical and syntactic features of ESL writing by students at different levels of L2 proficiency. *TESOL Quarterly*, 28(2), 414–420. <https://doi.org/10.2307/3587446>
- Ferris, D., & Hedgcock, J. (2014). *Teaching L2 composition: Purpose, process, and practice* (3rd ed.). Routledge.
- Gates, G., Cox, T. L., Bell, T. R., & Eggington, W. (2020). Line, please? An analysis of the rehearsed speech characteristics of native Korean speakers on the English Oral Proficiency Interview—Computer (OPIc). *Language Testing in Asia*, 10(1), 1–20. <https://doi.org/10.1186/s40468-020-00110-5>
- Green, A. (2007). Washback to learning outcomes: A comparative study of IELTS preparation and university pre-session language courses. *Assessment in Education*, 14(1), 75–97. <https://doi.org/10.1080/09695940701272880>
- Grabe, W., & Kaplan, R. B. (1996). *Theory and practice of writing. An applied linguistic perspective*. Longman.
- Guo, L., Crossley, S. A., & McNamara, D. S. (2013). Predicting human judgments of essay quality in both integrated and independent second language writing samples: A comparison study. *Assessing Writing*, 18(3), 218–238. <https://doi.org/10.1016/j.asw.2013.05.002>
- Gu, Q., & Brooks, J. (2008). Beyond the accusation of plagiarism. *System*, 36(3), 337–352. <https://doi.org/10.1016/j.system.2008.01.004>
- Hackers. (2021, September, 14). [PEET TOEFIC] Study PEET TOEIC at Hackers! Special Deals with 15% Discount. *Hackers Blog*. <https://blog.naver.com/hacademia/222500460239>
- Institute of International Education. (2022, Jan). *2021 Project Atlas infographics*. IIE. <https://www.iie.org/research-and-insights/project-atlas/explore-data/united-states>

- Jarvis, S., Grant, L., Bikowski, D., & Ferris, D. (2003). Exploring multiple profiles of highly rated learner compositions. *Journal of Second Language Writing, 12*(4), 377–403. <https://doi.org/10.1016/j.jslw.2003.09.001>
- Jeon, J. (2010). Issues for English tests and assessments: A view from Korea. *Language Assessment in Asia: Local, regional or global, 4*, 53–77. [https://koreatesol.org/sites/default/files/pdf/\(2\)%20Jeon-A%20View%20from%20Korea%20\(2010%20AsiaTEFL%20Book%20Series\).pdf](https://koreatesol.org/sites/default/files/pdf/(2)%20Jeon-A%20View%20from%20Korea%20(2010%20AsiaTEFL%20Book%20Series).pdf)
- Jo, H. (2020, June, 25). “Bermuda triangle” in Jongno? Language school big 3 is transforming. *NEWS1*. <https://www.news1.kr/articles/?3974267>
- Johnstone, B. (2017). Discourse structure: Parts and sequences. In B. Johnstone (Ed.), *Discourse analysis* (pp. 339–473). John Wiley & Sons.
- Kaplan, R. B. (2005). Contrastive rhetoric. In R. B. Kaplan (Ed.), *Handbook of research in second language teaching and learning* (pp. 399–416). Routledge.
- Kim, E.-Y. J. (2017). The TOEFL iBT writing: Korean students’ perceptions of the TOEFL iBT writing test. *Assessing Writing, 33*, 1–11. <https://doi.org/10.1016/j.asw.2017.02.001>
- Kim, S. (2021). Prepping for the TOEFL iBT Writing test, Gangnam style. *Assessing Writing, 49*, 100544. <https://doi.org/10.1016/j.asw.2021.100544>
- Knoch, U., Macqueen, S., & O’Hagan, S. (2014). An investigation of the effect of task type on the discourse produced by students at various score levels in the TOEFL iBT® writing test. *ETS Research Report Series, 2014*(2), 1–74. <https://doi.org/10.1002/ets2.12038>
- Knoch, U., Huisman, A., Elder, C., Kong, X., & McKenna, A. (2020). Drawing on repeat test takers to study test preparation practices and their links to score gains. *Language Testing, 37*(4), 550–572. <https://doi.org/10.1177/0265532220927407>
- Kyle, K. (2020). The relationship between features of source text use and integrated writing quality. *Assessing Writing, 45*, 100467. <https://doi.org/10.1016/j.asw.2020.100467>
- Kyle, K., & Crossley, S. (2016). The relationship between lexical sophistication and independent and source-based writing. *Journal of Second Language Writing, 34*, 12–24. <https://doi.org/10.1016/j.jslw.2016.10.003>
- Lee, I. (2018, Nov). English learning environment in mobile internet era. *Newsfreezone*. <http://www.newsfreezone.co.kr/news/articleView.html?idxno=88813>
- Liu, O. L. (2014). Investigating the relationship between test preparation and TOEFL iBT® performance. *ETS Research Report Series, 2014*(2), 1–13. <https://doi.org/10.1002/ets2.12016>
- Llosa, L., & Malone, M. E. (2019). Comparability of students’ writing performance on TOEFL iBT and in required university writing courses. *Language Testing, 36*(2), 235–263. <https://doi.org/10.1177/0265532218763456>
- Ma, J., & Cheng, L. (2015). Chinese students’ perceptions of the value of test preparation courses for the TOEFL iBT: Merit, worth, and significance. *TESL Canada Journal, 33*(1), 58–79. <https://doi.org/10.18806/tesl.v33i1.1227>
- Malone, M. E., & Montee, M. (2014). Stakeholders’ beliefs about the TOEFL iBT® test as a measure of academic language ability. *ETS Research Report Series, 2014*(2), 1–51. <https://doi.org/10.1002/ets2.12039>
- Messick, S. (1982). Issues of effectiveness and equity in the coaching controversy: Implications for educational and testing practice. *Educational Psychologist, 17*(2), 67–91. <https://doi.org/10.1080/00461528209529246>
- Park, J. (2012). Hackers stole the TOEIC test, the trick was also “cunning.” *NEWSIS*. <https://n.news.naver.com/mnews/article/003/0004326287?sid=102>
- Plakans, L., & Gebril, A. (2013). Using multiple texts in an integrated writing assessment: Source text use as a predictor of score. *Journal of Second Language Writing, 22*(3), 217–230. <https://doi.org/10.1016/j.jslw.2013.02.003>

- Riazi, A. M. (2016). Comparing writing performance in TOEFL-iBT and academic assignments: An exploration of textual features. *Assessing Writing*, 28, 15–27. <https://doi.org/10.1016/j.asw.2016.02.001>
- Roberts, M. (2002). TOEFL preparation: What are our Korean students doing and why. *The Korea TESOL Journal*, 5(1), 81–106. [https://koreatesol.org/sites/default/files/pdf\\_publications/KTJ5-2002web.pdf#page=89](https://koreatesol.org/sites/default/files/pdf_publications/KTJ5-2002web.pdf#page=89)
- Ross, S. J. (2008). Language testing in Asia: Evolution, innovation, and policy challenges. *Language Testing*, 25(1), 5–13. <https://doi.org/10.1177/0265532207083741>
- Schoonen, R., Van Gelderen, A., De Glopper, K., Hulstijn, J., Simis, A., Snellings, P., et al. (2003). First language and second language writing: The role of linguistic knowledge, speed of processing, and metacognitive knowledge. *Language Learning*, 53, 165–202. <https://doi.org/10.1111/1467-9922.00213>
- Singh, P. (2022, Jan). How will e-learning transform education in a post-COVID 19 world? *Training*. <https://trainingmag.com/how-will-e-learning-transform-education-in-a-post-covid-19-world/>
- Song, X., & Cheng, L. (2006). Language learner strategy use and test performance of Chinese learners of English. *Language Assessment Quarterly*, 3(3), 243–266. [https://doi.org/10.1207/s15434311laq0303\\_2](https://doi.org/10.1207/s15434311laq0303_2)
- Taylor, C., & Angelis, P. (2008). The evolution of the TOEFL. In C. A. Chapelle, M. K. Enright, & J. M. Jamieson (Eds.). *Building a validity argument for the TOEFL iBT* (pp. 27–54). New York, NY: Taylor and Francis.
- Trapman, M., van Gelderen, A., van Schooten, E., & Hulstijn, J. (2018). Writing proficiency level and writing development of low-achieving adolescents: the roles of linguistic knowledge, fluency, and metacognitive knowledge. *Reading and Writing*, 31(4), 893–926. <https://doi.org/10.1007/s11145-018-9818-9>
- Van Dijk, T. A. (1981). Episodes as units of discourse analysis. In D. Tannen (Ed.), *Analyzing discourse: Text and talk* (pp. 177–195). Georgetown University Press.
- Winke, P., & Lim, H. (2017). The effects of test preparation on second-language listening test performance. *Language Assessment Quarterly*, 14(4), 380–397. <https://doi.org/10.1080/15434303.2017.1399396>
- Yu, G., He, L., Rea-Dickins, P., Kiely, R., Lu, Y., Zhang, J., Zhang, Y., Xu, S., & Fang, L. (2017). Preparing for the speaking tasks of the TOEFL iBT® test: An investigation of the journeys of Chinese test takers. *ETS Research Report Series*, 2017(1), 1–59. <https://doi.org/10.1002/ets2.12145>
- Xie, Q. (2013). Does test preparation work? Implications for score validity. *Language Assessment Quarterly*, 10(2), 196–218. <https://doi.org/10.1080/15434303.2012.721423>
- Zhang, Y. (2008). *Repeater analyses for TOEFL iBT*. ETS. <https://www.ets.org/Media/Research/pdf/RM-08-05.pdf>
- Zhang-Wu, Q., & Brisk, M. E. (2021). “I must have taken a fake TOEFL!”: Rethinking linguistically responsive instruction through the eyes of Chinese international freshmen. *TESOL Quarterly*, 55(4), 1136–1161. <https://doi.org/10.1002/tesq.3077>
- Zheng, C., & Park, T. J. (2013). An analysis of errors in English writing made by Chinese and Korean university students. *Theory & Practice in Language Studies*, 3(8).

## Appendix

### Example Writing Templates

#### *Template Example 1 transcribed from Hackers Intermediate*

**Some people think that** people should try out several different jobs before settling into long-term employment. **However, in my opinion,** people should pursue one career without trying out several other jobs first. **This is because** it is helpful to the workers as well as to the companies that hire them. The workers can be more successful professionals faster, and the companies do not need to waste money.

**To begin with,** by making a commitment to a single field, people can become successful more quickly in their particular field...

**On top of that,** if people commit to a job for the long term, it will be financially helpful to their companies...

**To sum up,** if people choose a long-term carrier instead of trying out multiple jobs first, both the worker and the companies that hire them will benefit...

#### *Template Example 2 transcribed from Pagoda Actual Test*

**It is of great worth to discuss whether or not** teachers should be paid the same base pay as other professionals like doctors and lawyers. **The answer to this question can vary depending on each individual's stance. From my perspective, I agree/disagree with the given notion because of several reasons.**

**First of all,** becoming a lawyer and doctor require much more time and effort as being a teacher. **This is attributed to the fact that** before students become a doctor or a lawyer they need to handle various academic matters at the same time with a great pace...

**It is obvious that** lawyers and doctors spend a substantial amount of time at... **According to research conducted by DS, one of the most well-known research centers in South Korea,** teachers only work about 180 days per year with three months of vacation. **The same study also noted that** because of their shorter workdays, teachers actually earn more...

**As this demonstrates,** it does not make sense that teachers should be paid as much as...

**To conclude what I mentioned above, I disagree that** teachers should be paid the same base pay as other professionals like doctors and lawyers since in order to...