

論文

An Analysis of the Writing Performance of In-Service Teachers of English Participating in an Online Writing Course: Insights from Rubric Ratings

オンライン講座における現職英語教員のライティングスキルの分析：
ルーブリック評価からの洞察

Peter PARISE

ABSTRACT

本研究は、英語教員対象のオンライン講座「英文ライティング添削講座」で得られた、さまざまなデータを分析したものである。受講者の英作文の評価には IELTS (International English Testing System) のルーブリックを使用し、トータルスコアと各評価項目（課題への回答、一貫性とまとまり、語彙力、文法知識と正確さ）のスコアについて、分散と相関を調べた。その結果、受講者にとって慣れ親しんだライティングタスクより複雑なタスクの方が、ライティングスキルに顕著な差が見られることがわかった。相関分析の結果からは、タスクが複雑になるほどライティングのさまざまな側面に受講者の注意が向くことが推察された。これらのことから、英語教員のためのライティング指導では、段階的にタスクの難易度を上げながら、さまざまなトピックのタスクに取り組ませる必要がある、ということが示唆された。

INTRODUCTION

The intention of this paper is to analyze data related to the Support Writing Course (英文ライティング添削講座) for the 2018 and 2019 terms. This course provides in-service Japanese teachers of English opportunities to practice writing in academic English. This is vital for teachers since most of their experience of writing as language learners in junior high and high school focused mostly on translating prefabricated Japanese sentences to English (Reichelt, 2009, p. 197) and that their experiences as learners can have a significant impact on their teaching practice (Borg, 2003). Providing this opportunity for teachers to write in English and get vital feedback for it is necessary to give them confidence and experience to conduct writing in their own classes which tends to be a neglected skill as reported in qualitative studies of teacher practice (Casanave, 2009; Sasajima et al., 2012). In addition, the participants are also exposed to methods for teaching writing and one of these methods is the use of rubrics. This type of evaluation method is the focus of this paper, because it allows researchers to evaluate the writing skills and needs of learners based on discrete categories. Another vital aspect is that this is a blended online course. What this means is that about 30% of the course is face to face, but the remaining 70% is administered online (Boettcher & Conrad, 2010). This is beneficial to the participants because this allows them

to work at their own pace, at anytime, anywhere to practice writing. To provide this course online, the institute established a Moodle website to distribute content related to the course as well as enable learners to upload and download drafts with ease to be graded for their writing.

This paper will focus on the writing performance of the participants of the 2018 and 2019 terms with each term containing 2 cohorts of participants for both the spring and fall. The adoption of rubrics provides relevant feedback to participants but also helps to understand their needs as learners. One question that can be asked based on this data is: how do the ratings of the participants' writing reflect their performance for different types of tasks and in different skill dimensions to reveal what their needs are in terms of academic writing?

LITERATURE REVIEW

Rubrics as a Means for Evaluating Productive Skills

Rubrics are necessary for evaluating productive skills such as speaking and writing because they are consistent, analytical instruments. Rubrics have become more relevant to teachers of English in Japan due to policies advocated by MEXT which suggests English courses for high school focus on integrating the four skills (Tahira, 2012; 文部科学省, 2018) rather than divide them separately in skill specific courses. Due to these trends, the Support Writing Course adopted rubrics to not only evaluate the participants but also through the experience of this course, show them how a rubric can be used to evaluate writing for their own teaching practice.

Defining Rubrics

A *rubric* or *rating scale* is defined as an “established means of guiding raters to improve their level of agreement with their colleagues.” (Green & Hawkey, 2012, p. 301) Rubrics are valuable as an evaluation tool for writing and make explicit what is needed to achieve a task based on the criteria of each skill category detailed in the rubric. The essential parts of a rubric are as follows in Figure 1. The design of a rubric needs four specific things: the description of the task, the *dimensions* which are the vital skills to be demonstrated by the participant and *descriptions* of each level of the performance is associated with a specific *scale* or score. Appendix A shows the breakdown of the rubric used for this writing course. The value of rubrics can be appreciated from two perspectives. From the perspective of the instructor, a rubric is a consistent, standardized set of criteria that two or more instructors must agree upon to evaluate the performance of the learners. This agreement is measured with inter-rater reliability statistics, depending on the number of raters who are available (Hallgren, 2012).

- The task description.
- The scale or scores defining the "levels of achievement."
- The dimensions of the assignment.
 - "A breakdown of the skills/ knowledge involved in the assignment"
- The descriptions of each level of performance.
 - "Specific feedback" for the learner.

Figure 1. Important elements of a rubric from Stevens and Levi, (2013, p 6.).

Rubrics are necessary since they are a useful way to maintain fairness when scoring open-ended responses (Green & Hawkey, 2012, pp. 299–306). Rubrics must communicate clearly and explicitly the expectations of the instructor and the task itself to the learner (Hyland, 2003, p. 221). While doing the task, the learner can then be mindful of the dimensions and descriptions described in the rubric. Each descriptor of the rubric defines a certain level of ability and a corresponding score. Rubrics are essential as means for feedback by allowing the instructor to respond in a timely manner (Stevens & Levi, 2013, pp. 17–22). This is because the instructor can tie their responses to the descriptions and dimensions of the rubric which are already decided prior to the task, reducing the time needed for grading writing. For learners, the rubric can be used to self-assess performance (Dawson, 2017, p. 355) by showing them what is needed for improvement.

Rubrics also provide a way to track student progress. Analytic rubrics, depending on the complexity of the descriptors, can provide vital information about the skill of the writer (Casanave, 2017, p. 239). While the rubric in use is to provide feedback for learners, it can also be utilized to obtain a general view of how learners perform when engaging in writing. By tracking learner performance an instructor and researcher can determine how learners fare when engaging in L2 writing tasks. Research in L2 writing focuses on task effects, the influence of context where writing occurs, the process of writing, as well as the influence of instruction on how writing is performed. To gain insights into these diverse influences and conditions, a direct analysis of the writers' texts is warranted (Polio & Friedman, 2017, pp. 15–19). Holistic and analytic rating scales are regarded as methods to obtain data to evaluate text quality (Polio, 2012, p. 151). What rubric ratings allow is the ability to see general trends, which are evident in a group of learners rather than individual performances and allow teachers to formulate strategies for teaching writing as well as inform research. A collection of ratings from a standardized rubric can be collected to find trends in student performance, which can inform teachers about the current state of their instruction, and "help refine teaching methods"(Stevens & Levi, 2013, p. 28). In addition, it can also serve as a standard to evaluate the performance of learners in certain tasks. This is vital to

determine if the participants grasp the essential elements of academic writing in terms of content, organization, and their ability to use appropriate grammar and vocabulary. When looking at the development of learners' writing though, to believe that their progression follows a linear pattern is a faulty assumption because success in writing can vary depending on the task as well as the context (Slomp, 2012). However, a consistent rubric can reveal insights about the learners' responses and needs when engaging in certain writing tasks. A learner may be successful with a familiar topic such as a self-introduction paragraph, but their writing skills may be challenged with a less familiar topic such as summarizing an article and stating their opinion about it. What can be understood with a rubric is the needs learners have after they engage in certain writing tasks,

EFL Writing Needs in the Japanese Context

As mentioned in the introduction, teachers need writing practice and guidance in writing in an academic manner because their experience with writing has been limited due to the emphasis on grammar translation. Since describing every detail of the errors performed by Japanese L2 learners of English is beyond the scope of this article, this section will focus on some generalized aspects of the needs learners have when writing academically in English. In terms of language-focused writing issues that are most common, Japanese learners are observed to experience confusion with subject-verb agreement, due to the topic prominent aspect of Japanese, as well as how and where to place prepositions and coordinating conjunctions (Celce-Murcia & Larsen-Freeman, 1998). In terms of vocabulary related issues, one issue that not only impacts Japanese learners but also most L2 writers of any language is the confusion in word choice. L2 writers tend to choose words that are more appropriate for spoken rather than written and academic registers (Gilquin & Paquot, 2007, 2008). In regards to content and organization issues for writing early contrastive rhetoric studies claimed that L2 writers from Asian cultures were believed to follow certain patterns in contrast with writers of Western cultures and prefer more inductive and "reader responsible" approaches to text organization (Hinds, 1983; Kaplan, 1966). However, this position has been challenged in studies that have demonstrated that Japanese learners do have the ability to structure their texts in a more deductive way (Hirose, 2003; Stapleton, 2002). This issue may be grounded in the past learning experiences of Japanese students when learning to write in their L1, which tends to focus more on connecting with readers emotionally (Hirose, 2003, p. 183). Japanese university learners not displaying "critical thinking" in their writing may not be a result of the cultural norms of Japan, but that the L1 writing tasks for students in general education focus more expressing their "personal" outlook and instructed to structure their texts in a freer, stream-of-consciousness form. This is shown in the writing of *kansobun*, which is a common writing task for students when sharing specific experiences (McKinley, 2013, p. 11).

Considering these diverse needs, this article intends to report on the writing performance of Japanese teachers of English. Since most of the research in EFL writing in general is focused on university students, investigating this population of language learners who are at the same time language-teaching professionals can yield interesting insights. This article hopes to show what teachers need in terms of in-service training and guide any future policy decisions on the necessity of this type of training. The chosen means for determining these needs is to use the IELTS rubric to match the expectations of academic writing in English with their performance on four writing tasks to establish general trends.

Based on the literature reviewed addressed above, this article poses the following research questions.

- 1 What can the results of an analysis of the IELTS rubric ratings for four different writing tasks as well as between the ratings of different skill dimensions between these tasks reveal about the participants' academic writing needs?
- 2 Based on the insights obtained from research question 1, what kind of interventions are warranted to address the academic writing needs of Japanese in-service teachers of English?

METHOD

Participants

The participants of this study consist of two cohorts of the Support Writing Program from the 2018 and 2019 terms. Each year consists of a spring and fall semester, with approximately 10 or more participants for each grouping. At the beginning of each course, the participants filled in questionnaire that inquired about their background, including teaching experience, and performance on proficiency tests. Table 1 below shows the traits of the applicants for both years who responded to the background questionnaire.

Table 1 *Counts of Traits the Applicants of the Writing Program of 2018 and 2019 (N = 34)*

Gender	Age Range		School		Years Teaching		
Female	25	20 – 30 years	7	High School	30	1-5 years	8
Male	9	30 – 40 years	11	Junior High	4	5 – 10 years	15
		40 – 50 years	11			10 – 20 years	2
		50 – 60 years	5			20 + years	2

What we can know from these counts is that 74% of the participants are female with the remaining 26% are male. The age range shows most of the participants, about 64% range from 30 to 50 years of age with 67% having 1 to 10 years of experience teaching English. Most of the participants of this program are from the public high school context. In addition to these counts

the participants also reported the type of English proficiency test that they have taken in the past and the means and standard deviations of the scores for the respondents are shown in Table 2.

Table 2 *Counts and Descriptive Statistics of Participants' Test Experience (N = 34)*

Test Taken	TOEIC	TOEFL	IELTS
Yes	29	13	8
No	4	20	25
Mean (SD)	852 (359)	86 (40)	7(3)

Note: All TOEFL scores were converted to TOEFL iBT scores for the sake of comparison.

The participants of this study are those who have successfully completed the course by completing four writing assignments as well as the quizzes on Moodle, which check their understanding of the content of the course. The breakdown of successful candidates of the writing course is shown in Table 3. The accumulated number of successful participants from these four cohorts is 30 individuals who are in-service teachers of English in both junior and senior high schools for the Kanagawa prefecture.

Table 1 *Number of the Participants and Successful Candidates in the 2018 and 2019 Cohorts.*

Cohort	Spring 2018	Fall 2018	Spring 2019	Fall 2019
# of participants	9	11	10	11
# successful	6	8	10	6
% success	66.7	72.7	100	54.6

The purpose of examining only the successful participants is because a full set of scores from each assignment is necessary to obtain a clearer view of their needs in terms of academic writing. Table 4 shows basic descriptive stats noting the mode age, years teaching, as well as mean TOEIC score between each cohort between Spring 2018 to Fall 2019.

Table 4 *Descriptive statistics regarding the participants in the 2018 and 2019 Cohorts.*

Cohort	Spring 2018	Fall 2018	Spring 2019	Fall 2019
Female to Male Ratio	7:2	8:3	3:2	4:0
Mode Age	20-30	40-50	30-40	30-40
HS/JHS Ratio	7:2	10:1	9:1	4:0
Mode Years Teach	1-5	5-10 10-20	5-10	5-10
Mean TOEIC	814	870	878	863

*Note: There were only four out of eleven participants responded to the background survey from the Fall 2019 cohort.

Context

Moodle

The context of this course as discussed in the introduction is a blended online environment, which means that part of the course is conducted in a face-to-face session at the beginning and at the end of the course, but the middle is entirely online. The online portion is managed using Moodle, an open source learning management system (LMS) which allows for the delivery and evaluation of content produced by the learners and the instructors. Moodle allows teachers to build a website for learning based on constructionist principles and creates a space for “connected knowing” to facilitate “productive educational relationships” (Dougiamas & Taylor, 2003, p. 2). Moodle supports learning by providing learners the ability to gauge their own learning with gamification elements such as badges and the progress bar (de Raadt, 2014; Greeve, 2014; Parise, 2015, 2019).

The Moodle website is organized in modules that are based on the four assignments offered in the writing course, plus two modules dedicated to the two face-to-face sessions as well as five additional modules to support the participants’ writing development. The four modules that focus on the assignments contain materials, forms, as well as a quiz to check the participants understanding of the content.

Assignments in the Writing Program

There are four main assignments offered in the writing course that are staged in a progression of topics which may be familiar to the participants but gradually becoming more complex as the course progresses. The breakdown of the assignments are as follows in Table 5. The description of the assignments shows a variety of task demands which can be useful when making connections between the rubric ratings and the conditions of the task. One task demand is a gradually increasing word count. This condition encourages the participant to generate more words to fulfill the task and may reveal their needs in terms of language related aspects such as grammar and vocabulary. Other demands are the need to use conjunctions and adjectives for the participants to describe themselves fully for the first assignment and in a more specific fashion challenge them in terms of word choice and order.

Content related demands are those such as the use of discourse markers in order to organize the participants’ writing in a logical manner, as well as the need to give examples that the reader can “see or imagine” are demands that call for a need give a clear description. These demands, while specific to some writing assignments, are common in others too. In order to make sense of the results of the ratings analysis, these task demands must be referenced in order to obtain a clearer view of the needs of in-service Japanese teachers of English in terms of their skill as writers of academic English.

Table 5 Topics and Task Demands for the Support Writing Course (英文ライティング添削講座)

Assignment	Task Demands
Writing about a Person	Participants must: <ul style="list-style-type: none">• Use at least three conjunctions• Use three physical adjectives• Use three personality adjectives• Write at least 130 words but no more than 200 words.• Make adequate use of transitions.
Writing about Your Opinion	Participants must: <ul style="list-style-type: none">• Use the rule of three at least once• Give examples that the reader can see or imagine.• Use a counter argument in the second reason.• Summarize their points and state their opinion again in the concluding sentence.• Write at least 200 words but no more than 300 words.• Make adequate use of transitions.
Writing to Compare and Contrast Two Things	Participants must: <ul style="list-style-type: none">• Use at least 3 signal words for comparison or contrast• Give examples that the reader can see or imagine.• Write at least 250 words.• Make adequate use of transitions.
Writing to Paraphrase a News Article and Give Your Opinion	Participants must: <ul style="list-style-type: none">• Summarize the main ideas of the article by answering the 5W-1H questions.• Give examples to support the main point.• Support their opinion regarding the topic using some data or statistics.• Write at least 300 words.

Instruments

Google Forms

A Google Forms was used to conduct a survey of the participants of these courses. This instrument was used to gather background data, allowing us to gauge the gender, age range, years of teaching, as well as which type of tests they engaged in to get full picture of the participants language learning experience. The survey was administered in the first face-to-face class meeting via the institute's computers. The responses were converted to spreadsheet form via Google Drive and analyzed to obtain the measures reported in the participant section of this article.

Adapted IELTS Rubric

The IELTS (International English Language Testing System) rubric is used to evaluate academic writing performance for its four skills focused test and “measures the language proficiency of people who want to study or work where English is used as a language of communication.” (IELTS, 2020a). The writing section of the test measures the academic writing proficiency of those intending to attend undergraduate or postgraduate studies, or to attain a professional position abroad (IELTS, 2020b). The rubric utilized for the course is adopted from the IELTS Task 2 grading rubric for evaluating the writing portion of the test. The adapted rubric uses the four categories: Task Achievement, Coherence and Cohesion, Lexical Resource, and Grammatical Range and Accuracy. The first two categories describe the degree of which the participant addresses the task and maintains the logical progression and structure of the text. The latter categories focus on the language aspect of the assignment, in other words how the participant uses their inter-language to build a text. The benefit of using IELTS is its accessibility. The rubric is available online. In addition, the four skill categories are comprehensible for learners and should a teacher adopt writing in their own courses, they can use similar skill dimensions for their own rubrics based on the IELTS design. However, the version used for this study is an adapted version of the official rubric used for evaluating the test. For the writing course, the categories have been reduced from the initial nine grades to five. One reason for doing this is to reduce the complexity for the participants and enable them to look at the rubric as a form of feedback. Nine grade categories may be fine for evaluators but for the participants, the descriptors between each grade is too subtle to use as a self-evaluation tool. This may be especially true for teachers who are unfamiliar with rubrics.

Procedure

Rubric ratings are given for each draft submitted by the participants. The participant uses a ready-made form which includes a variety of feedback options, with the rubric available near the end of the form. After two instructors look at the draft, make corrections and comments, one instructor gives scores for each category of the rubric and then totals the ratings to become a total score for the draft. If there is any disagreement regarding the rubric scoring, then the instructors discuss the issue to agree on the final scores.

The draft ratings are the main source of data for this paper. Scores for final drafts were excluded from this study. One reason for this choice is that the final drafts are essentially the participants responses to feedback from the instructors, and any changes done for this draft does not truly reflect their writing ability. The final drafts entail the participants’ repair of sections of their writing related to either errors in content, errors in word choice, or accuracy. In addition, the final draft is not starting from a blank slate, where the participants are drawing from their writing

ability. Hence, this study chooses to focus only on the first drafts to measure the participants' writing performance.

Analysis

This study will compare the rubric ratings of the drafts of four assignments of the 30 successful candidates. These consist of each category of the rubric to make comparisons. In addition, the final scores of the drafts will also be taken into consideration to look for any larger trends between the participants as a whole and between cohorts. These four different assignments were taken in linear order from the beginning of the course until the end and evaluated with one consistent rubric. This necessitates the use of statistical analysis to determine if there are significant differences between assignments when comparing cohorts and the whole population over this time period. The breakdown of the data sets are as follows:

- Draft total scores- the sum of all the rubric dimension scores culminating in one final score for the draft.
- Dimension scores- comparisons of dimensions between assignments.
 - Task Achievement
 - Coherence and Cohesion
 - Lexical Resource
 - Grammatical Range and Accuracy

By looking at these sets of data, we can understand where the participants experienced any difficulty based on their writing performance for different tasks.

A repeated measures analysis of variance (ANOVA) will be utilized to determine if there are any significant differences between cohorts in relation to draft scores on the tasks, as well as individual dimensions measured on the rubric. In addition, the participants will also be compared to see if there are any significant differences between writing tasks. One of the main assumptions that must be met to conduct this analysis is if the data is normally distributed, in other words the data must be distributed in the form of a bell curve. To check whether this data is normal or not, Shapiro Wilk's Test for Normality will be conducted on all the data sets in this study. If the measures are significant, then it is assumed that the scores are not a normal distribution. Friedman's Rank Sum Test a non-parametric type of ANOVA will also be utilized. This type of statistic does not rely on normally distributed data will be used to determine if there are significant differences between each dimension of the rubric to gauge how the participants perform as they engage in four different writing tasks.

An additional analysis will be the use of correlations between the final draft scores and the dimension scores for each rubric. A non-parametric statistic Kendall Tau will be used to determine the degree of relationship between certain dimension scores and the final draft.

Histograms will also be used to make visual comparisons of the data. These are to compliment the statistical measures and help to make immediate comparisons. All these analyses and graphics will be generated using R (R Core Team, 2020) via Rstudio (RStudio Team, 2019).

RESULTS

Draft Total Scores between Writing Tasks

Before conducting a repeated measures ANOVA on the data, a Shapiro Wilk's test for normality was conducted on the data. This yielded a non-significant result, which means that the data, the total scores of each draft of the four assignments for the Support Writing Course are normally distributed. The results of the repeated measures ANOVA found no significant differences between cohort $F(3,9) = 0.87, p > .05$ and participant $F(29,87) = 0.81, p > .05$. While the differences between scores for each writing assignment is not significantly different, this result shows that this population of teachers is equal between the different cohorts and the individuals in the courses themselves, and so the means of these four assignment scores represent one consistent group. All these distributions are negatively skewed, which means that the majority of the scores by these participants tend to be in the higher point ranges above the mean implying that from the first draft, most participants demonstrated that they possess good skills in writing in English as shown in Table 6. However, while the means for each assignment are relatively consistent, the standard deviations between the scores on assignments increased to almost 2.50. This is also supported by the negative kurtosis measures which means that the distribution of scores is collected more in the center, with fewer scores at the extremes, and is flatter (Field et al., 2012, p. 21).

What this implies is that while there is variation between the scores in these assignments, the participants total scores tended to gather at the mean. Looking at the histograms in Figure 2, for the draft of Assignment 1 there are apparently two peaks in this distribution, one at the mean itself and another two standard deviations above the mean. This implies that while some struggled with the assignment of writing about themselves in the third person, others performed it relatively well, to almost the top score. This two-hump distribution is also apparent for the Assignment 3 draft, about a quarter of the participants scored below the mean, implying that there was some difficulty with comparing and contrasting something in English while others seem used to the concept and performed well on their first draft. Since these are general scores of the first draft of these assignments, the dimension scores in the following sections will provide more details on the needs participants have with academic writing in English.

Table 6 Descriptive Stats of Total Draft Scores for Assignments 1-4 of the Support Writing Course

	Assignment 1 Draft	Assignment 2 Draft	Assignment 3 Draft	Assignment 4 Draft
Mean	15.60	15.33	15.67	15.73
S. Deviation	1.87	1.99	2.44	2.36
Skewness	-0.41	-0.02	-0.36	-0.25
Kurtosis	-0.90	-1.14	-1.16	-1.02

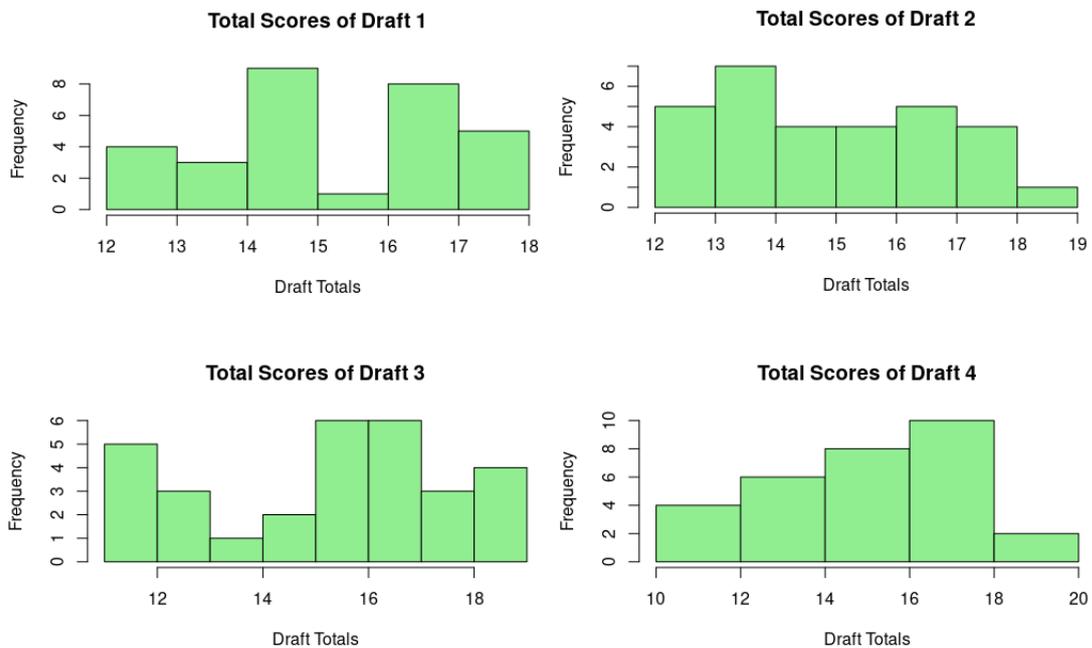


Figure 2. Histograms of the draft total scores for Assignments 1-4.

Dimension Scores between Writing Tasks

Task Achievement

The dimension of task achievement is a score related to the ability of the participants to complete the expectations of the task. *Task Achievement* is “the criterion under which the marker assesses the extent to which we have answered the question comprehensively and with relevant material.” (“The IELTS Writing Rubric,” 2019b). In other words, to what degree does the participant's response match the demands of the prompt? Regarding a writing task, does the participant address the question and the demands of the task, or do they diverge?

Table 7 Descriptive Stats of Task Achievement for Assignments 1-4 of the Support Writing Course

	Assignment 1 Draft	Assignment 2 Draft	Assignment 3 Draft	Assignment 4 Draft
Mean	3.93	3.83	3.97	3.90
S. Deviation	.74	.83	.72	.96
Skewness	.10	.30	.04	-.04
Kurtosis	-1.23	-1.55	-1.13	-1.54

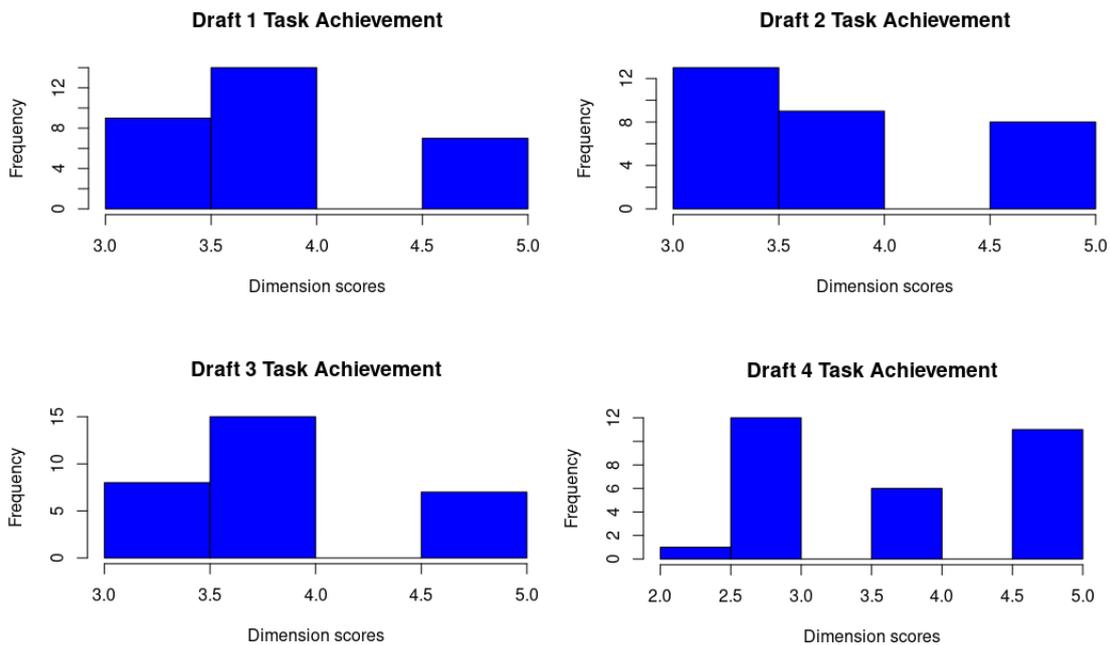


Figure 3. Histograms of the task achievement scores for Assignments 1-4.

The rubric as shown in Appendix A is a five-level scale with five being the highest and zero being the lowest. Before comparing these scores, a Shapiro Wilk’s analysis was conducted on this data to determine if the distribution is normal, therefore fulfilling the assumption of normality for using parametric statistics such as an analysis of variance as discussed in the previous section. The measures of these scores were significant, showing that this data is not a normal distribution. A non-parametric statistic Friedman’s Rank Sum Test was used, showing that there were no significant differences for the participants regarding these scores $\chi^2(3) = 0.80, p > .05$. The mean of these scores consistently neared a score of 4 implying that some of the participants in these courses were able to satisfy most of the expectations of the task for the assignments.

The standard deviation for all these assignments stayed slightly under .90 except for Assignment 4. This means that for Assignment 4, the ability for participants to meet the task demands were varied. Looking at the histograms in Figure 3, Assignment 4 has a wide range of scores with two humps forming, both at two extremes. Almost half of the participants were able to score 5 in their draft for Assignment 4, but another half were unable to score above 3. This assignment, which asks for the participants to summarize a news article in one paragraph and then share their opinion about it in the next, may have been a challenge for some. This may be due to the complexity of the assignment as well as a test of the participants’ skill of paraphrasing and

may reveal a deficiency in the participants understanding of what paraphrasing entails.

Coherence and Cohesion

Coherence and Cohesion can be defined as the organization of ideas, how those ideas are developed as well as the use of paragraphing, how those paragraphs are sequenced and if “cohesive devices” are utilized to “indicate the relationship between these paragraphs.” (“The IELTS Writing Rubric,” 2019a). Regarding this dimension a Shapiro Wilk’s test was conducted, which gave a significant result showing that this data is not normally distributed, and a Friedman's Rank Sum Test was conducted $\chi^2(3) = 0.5, p > .05$ yielding a non-significant result.

Here the average scores are consistently at around 4.0 in terms of scoring and have very consistent standard deviations for the first two assignments at around .80 and around .87 for the final two assignments as shown in Table 8. In Figure 4, we can see that assignments 3 and 4 are negatively skewed more than assignments 1 and 2, showing that while there is a majority of participants who can organize their draft in a coherent manner from the first draft, some seemed challenged by the task.

Assignments 3 and 4, while negatively skewed, show that half of the participants scored a 4 and above and reveals that most of the participants were able to organize their first draft well. However, for some this task can be conceptually more complex. Assignment 3 is a compare and contrast type task, and the latter Assignment 4 as mentioned in the previous section, is a task where the participant must summarize an online article and then give their opinion about it. These tasks involve more organization, and the challenge of these tasks lies in the planning stage of writing, where the participant must organize the their thoughts before writing either with an outline or with a graphic organizer such as the Venn diagram.

Lexical Resource

Lexical Resource is defined as the learner’s ability to use synonyms, collocations, and appropriate word families as well as accurate usage of basic vocabulary (*IELTS Writing Tips for Teachers - Lexical Resource | Take IELTS*, n.d.).

Table 8 *Descriptive Stats of Coherence and Cohesion for Assignments 1-4 of the Support Writing Course*

	Assignment 1 Draft	Assignment 2 Draft	Assignment 3 Draft	Assignment 4 Draft
Mean	4.10	4.00	4.06	4.16
S. Deviation	.80	.79	.86	.87
Skewness	-0.17	0.00	-0.43	-0.61
Kurtosis	-1.48	-1.44	-0.91	-0.75

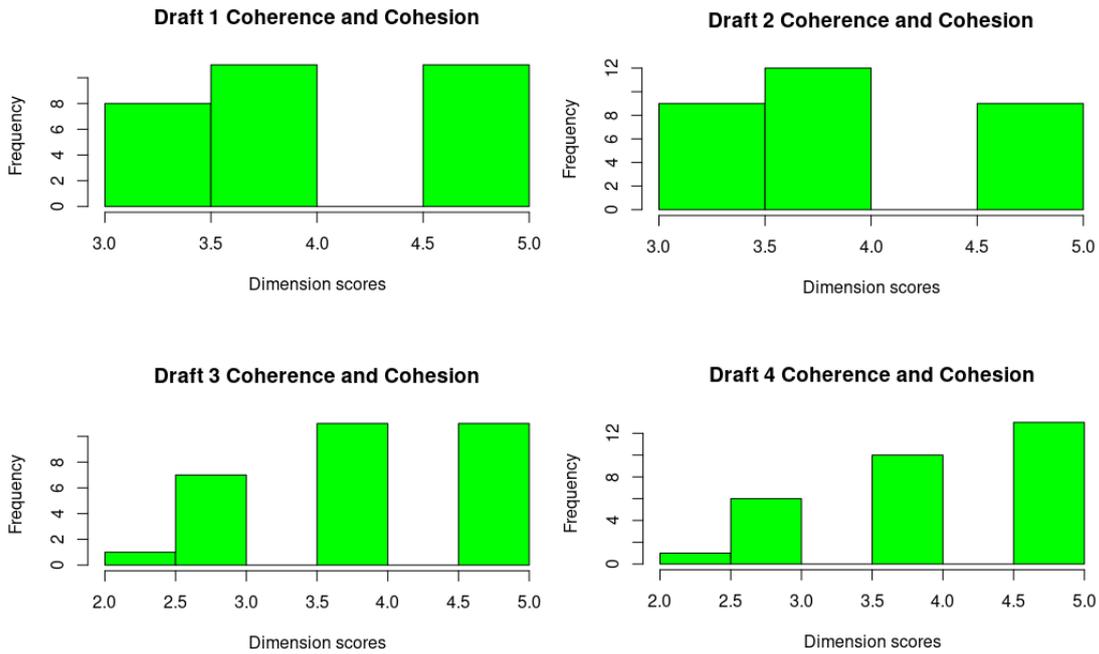


Figure 4. Histograms of the coherence and cohesion scores for Assignments 1-4.

This dimension also yielded a significant Shapiro Wilk’s result, and the Friedman's Rank Sum Test result was non-significant $X^2(3) = 0.84, p > .05$. The mean scores for this dimension are lower compared to the task achievement and coherence and cohesion dimension scores, however these are quite consistent between assignments with the mean for Assignment 1 and 2 being consistently at 3.77 and Assignment 3 and 4 being consistently at 3.83 as shown in Table 9. The standard deviations between these scores is at around .50 except for Assignment 3, clearly showing that there is little variation between the participants in terms of lexical resource. The histograms in Figure 5 show, while skewed negatively, that a majority of the participants rarely goes above a score of 4 in their initial draft, and this is marked especially in Assignment 1.

Table 9 Descriptive Stats of Lexical Resource for Assignments 1-4 of the Support Writing Course

	Assignment 1 Draft	Assignment 2 Draft	Assignment 3 Draft	Assignment 4 Draft
Mean	3.77	3.77	3.83	3.83
S. Deviation	.43	.50	.59	.46
Skewness	-1.20	-0.38	.04	-0.60
Kurtosis	-0.58	-0.34	-0.44	0.51

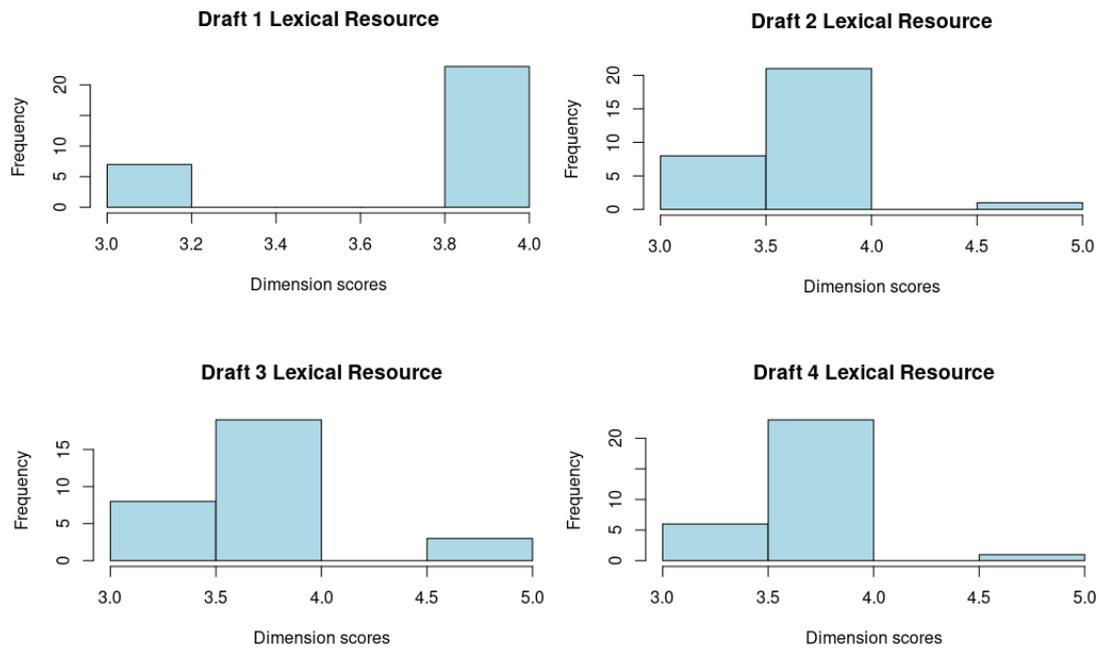


Figure 5. Histograms of the lexical resource scores for Assignments 1-4.

Each assignment in this course makes demands in terms of word choice, such as the use of different adjectives in Assignment 1 to describe himself or herself in the third person. However, a majority of the participants with a score of 4 shows that they do possess the vocabulary to write their drafts and the additional feedback to be given for their final draft is what they need to expand that vocabulary.

Grammatical Range and Accuracy

Grammatical Range and Accuracy is the test taker's ability to accurately use a variety of tenses, complex and simple sentences, and punctuation, all of which impact communicating meaning to the reader (*IELTS Writing Tips for Teachers - Grammatical Range and Accuracy | Take IELTS*, n.d.). This dimension was also analyzed with Shapiro Wilk's, obtained a significant result, and the Friedman's Rank Sum Test was used to determine if there were significant differences between the scores of each assignment. Like the other dimensions this too was non-significant $X^2(3) = 0.61, p > .05$.

Table 10 *Descriptive Stats of Grammatical Range and Accuracy for Assignments 1-4 of the Support Writing Course*

	Assignment 1 Draft	Assignment 2 Draft	Assignment 3 Draft	Assignment 4 Draft
Mean	3.80	3.70	3.83	3.80
S. Deviation	0.48	0.47	0.59	0.53
Skewness	-0.49	-0.83	.04	-0.17
Kurtosis	0.17	-1.35	-0.44	-0.02

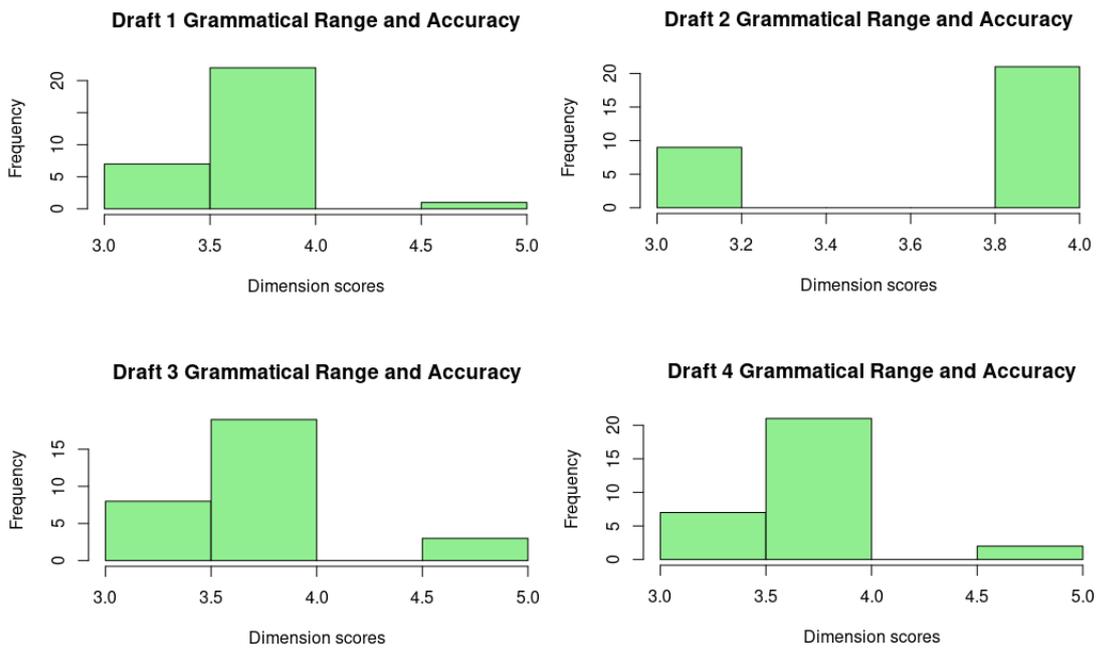


Figure 6. Histograms of grammatical range and accuracy scores for Assignments 1-4.

Like lexical resource, the mean of this dimension is less than 4 showing that these assignments are consistently difficult in terms of the accuracy and use of grammar. The standard deviation between these scores is consistent with Assignments 1 and 2, hovering around .48 and increasing slightly for Assignments 3 and 4, as shown in Table 10.

In terms of kurtosis, Assignment 1 is positive showing that most participants all collected at the rating of 4 and reveal that this assignment was consistently difficult as a task. Assignments 3 and 4, a small minority of participants were able to score a rating of 5, revealing a degree of familiarity with the language associated with these tasks. However, it should be noted that this is a very small minority of the participants of this study.

Correlations between Total Scores and the Dimensions

In addition to the analyses of variance and the descriptive statistics, correlations were

conducted to determine the relationship between certain skill dimensions with the total draft scores for each assignment. Since the prior Shapiro Wilk’s analysis in the previous sections yielded both significant and non-significant results, the best course of action when comparing these two data sets was to use the non-parametric correlation statistic Kendall Tau.

Table 11 *Correlations between the Assignment Draft Scores and the Dimension Scores*

	Task Achievement	Coherence and Cohesion	Lexical Resource	Grammatical Range and Accuracy
Writing about a Person	0.59	0.64	0.46	0.59
Writing about Your Opinion	0.46	0.47	0.55	0.58
Writing to Compare and Contrast Two Things	0.61	0.64	0.66	0.71
Writing to Paraphrase a News Article and Give Your Opinion	0.63	0.60	0.56	0.57

Note: All correlations are significant to the $p < .001$

The analysis resulted in all the dimensions to be significantly correlated with the total scores $p < .001$ and are shown in Table 11. The degree of relationship between these dimensions and the assignments are medium ($\pm .3$) to large ($\pm .5$) effect (Field et al., 2012, p. 209). With the first assignment, “Writing about a Person” we can see that there is a large effect for task achievement, coherence and cohesion and grammatical range and accuracy and a medium effect with lexical resource. This seems to imply that these skills may have been needed more from the participants, and vocabulary may have been less of a demand on the participants. In other words, while the tasks call for the participants to use physical and personality adjectives as mentioned in Table 5, the vocabulary they used may have been easier for the participants to generate because they were writing about themselves. However, organizing them in a text may have been slightly challenging due to the large effect on task achievement and coherence and cohesion scores and demanding for grammatical range and accuracy.

The correlations for the second assignment, “Writing about your Opinion” while the effect for grammatical range and accuracy is consistently large as the previous assignment, the degree of effect for the prior three: task achievement, coherence and cohesion, are of medium effect and lexical resource has a large effect in terms of degree of relationship. This may entail that the participants are quite familiar with the concept of expressing their opinion, but may be

challenged more in terms of vocabulary because they had to write about their opinion about an online article of their choice, but also had to provide examples which the reader can “see or imagine” as stated in Table 5. This may have facilitated more demands in terms of language related skills such as lexical resource and grammatical range and accuracy rather than the more content related skills.

The third assignment “Writing to Compare and Contrast” shows the largest effect with all the skills having a measure over $\pm .50$. What is notable is the degree of relationship between grammatical range and accuracy, which is at 0.71. This may imply that compare and contrast demands more from the participants in terms of how to organize ideas, but also how language related skills are vital as well to create a text that is logically structured. This overall large effect for these skills may mean that compare and contrast tasks require a wide range of skills from the participants for both content and language.

The final assignment, “Writing to Paraphrase a News Article and Give Your Opinion” like the previous assignment also obtained large effects across all skills, but not to the same degree as Assignment 3. The largest effect is for task achievement implying that the challenge for participants for this assignment was meeting the demands of the task. While the other skills also made demands on the participants, task achievement was mediated by the complexity of the task of paraphrasing and offering their opinion in two different paragraphs. The familiarity of the task of writing about one’s opinion, which was already performed in Assignment 2 is repeated again in the final assignment may have contributed to a large effect due to the complexity of the task. This is the same for coherence and cohesion. Initially in Assignment 2, this dimension was of medium effect, however with greater complexity, the effect was larger. This may be due to the combination of two skills, summarizing, which is the novel aspect of the assignment, paired with the already familiar task of sharing opinions.

Comparing all the correlations between these assignments for all these dimensions in Table 11, we can see a gradual rise in effect for all these skills. Task achievement tends to rise consistently but coherence and cohesion, lexical resource, and grammatical skill and accuracy gradually rise in effect until the third assignment, and fall with the fourth assignment. What this may mean is that the degree of familiarity with the topics of the first two and the last one has a relatively medium to slightly high effect on the participants, but the most challenging of all for them, and the most demanding in terms of skill is the compare and contrast assignment.

DISCUSSION

What can the results of a statistical analysis of the IELTS rubric ratings for four different writing tasks as well as between the ratings of different skill dimensions between these tasks

reveal about the participants academic writing needs? To answer this first research question, the results show that while there are no significant differences between these scores, some rather consistent needs are apparent for the whole group of participants examined. One is the issue of complexity of the assignments which show, depending on the dimension involved, whether the participants were challenged when writing their draft or not. Looking at the histograms of the total scores of the writing assignments, we can see that there are disparities within this group of participants in terms of split distributions and this is most notable for Assignments 1 and 3. Both these assignments make demands on the participants in terms of language use, such as using adjectives, and cohesion and coherence such as organizing a compare and contrast. One positive observation from these total scores is that the most similar tasks, Assignment 2 which is to state your opinion about a news article and Assignment 4 which is to summarize a news article and state your opinion about it seem to shift in terms of skew and kurtosis. This implies that the participants may have improved in this aspect of the course through repeated exposure to this type of task. In terms of the dimensions of the rubric, the most notable variation among the participants is in their task achievement and coherence and cohesion scores, when comparing histograms there is greater disparity in terms of scoring for the final two Assignments 3 and 4. These more complex tasks, as shown in the correlations and the analysis of descriptive statistics and the histograms, show that the final two assignments provided a positive challenge for the participants.

Based on the insights obtained from research question 1, what kind of interventions are warranted to address the academic writing needs of Japanese in-service teachers of English? For this second research question, while there are no significant differences between the assignments in terms of the total draft scores and the dimension scores one pattern is evident from the correlation results. Training teachers to write better starts with a balance of familiar topics, as demonstrated with the transition between Assignment 2 to Assignment 4, with more complexity such as compare and contrast and the addition of summarizing in the final assignment. These are relevant for addressing content related needs such as task achievement and coherence and cohesion, which deal directly with issues of how the logic of the text is structured. Given the opportunity to write and learn how to organize a text logically is a vital skill and should be an integral part of in-service teacher training. The scores with the most consistency appear to be the language related dimensions such as lexical resource and grammatical range and accuracy. These dimension scores seem to change gradually and can be addressed by offering consistent writing practice combined with relevant feedback to promote growth.

The results of this study however should be interpreted with caution. The fact that all the scores are non-significant shows that while the assignments are different, the participants performed in a very similar fashion. This begs the question whether these assignments are altogether different from each other. What is data does reveal is the ability of the participants over

the course of four assignments. However, means only report an average description of these participants' skills, and we can only get a glimpse of their needs in a general sense. A more detailed analysis is warranted which takes into account the wide diversity of the participants of this study. Adopting an item-response approach such as Rasch analysis can provide a more fine-grained view of the difficulty experienced by certain participants or cohorts. This opens doors for further research. Finally, inter-rater scores were not reported in this study. Only one instructor administered these scores as mentioned in the methods section and any discrepancy was discussed. However, this is not as reliable as an inter-rater score.

What can be certain from these rubric measures is that the more complex the assignment becomes, the greater disparity between participants as evident with the total scores of the first draft. This shows that giving teachers opportunities to tackle these types of writing assignments not only expose their needs but also provide them with the opportunity to grow as language learners. As evidenced from the correlations between the total draft scores and the dimension scores, the large effects on the dimension skills especially for the more complex tasks show that the teachers were challenged in a positive way.

CONCLUSION

The intention of this paper was to analyze data related to the Support Writing Course for the 2018 and 2019 terms. Using an adapted IELTS rubric and analyzing the total draft scores as well as the specific dimension scores, we are provided a snapshot of what in-service Japanese teachers of English in junior and senior high need in terms of academic writing development. What is understood that providing these teachers with writing assignments that make them to go outside their comfort zone by initially offering familiar assignments such as writing about themselves and their opinion, to more complex ones such as comparing and contrasting, to summarizing and adding their opinion, allow them to exercise content and language related skills of the IELTS rubric. In-service language teachers need training courses to expand their skills to meet the demands of education policy changes. However, to expect teachers to do this independently, considering their work and family obligations, without any organized teacher training course may be an impossible task for them. Instruction for in-service teachers for academic writing is a vital need that should not be neglected. March 2021 will mark the closing of the Kanagawa Prefectural Institute of Language and Culture studies, and with it, this writing course will end. The author hopes that in the future, some form of support for in-service language teachers in the Kanagawa Prefecture will be available to help them with their academic writing needs.

REFERENCES

- Boettcher, J. V., & Conrad, R.-M. (2010). *The online teaching survival guide: Simple and practical pedagogical tips* (1st ed). Jossey-Bass.
- Borg, S. (2003). Teacher cognition in language teaching: A review of research on what language teachers think, know, believe, and do. *Language Teaching*, 36(2), 81–109.
<https://doi.org/10.1017/S0261444803001903>
- Casanave, C. P. (2009). Training for writing or training for reality? Challenges facing EFL writing teachers and students in language teacher education programs. *Writing in Foreign Language Contexts: Learning, Teaching, and Research*, 256–277.
- Casanave, C. P. (2017). *Controversies in second language writing: Dilemmas and decisions in research and instruction* (Second edition). University of Michigan Press.
- Celce-Murcia, M., & Larsen-Freeman, D. (1998). *The Grammar Book: An ESL/EFL Teacher's Course, Second Edition* (2nd ed.). Heinle ELT.
- de Raadt, M. (2014, May 7). *The Progress Bar block*.
https://www.youtube.com/watch?v=06LA5Cv9Fhw&feature=youtube_gdata_player
- Dougiamas, M., & Taylor, P. (2003). Moodle: Using learning communities to create an open source course management system. *Proceedings of the EDMEDIA 2003 Conference*. EDMEDIA 2003, Honolulu, HI.
- Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. SAGE Publications Ltd.
- Gilquin, G., & Paquot, M. (2008). Too chatty: Learner academic writing and register variation. *English Text Construction*, 1(1), 41–61. <https://doi.org/10.1075/etc.1.1.05gil>
- Gilquin, G., & Paquot, M. (2007). Spoken features in learner academic writing: Identification, explanation and solution. *Proceedings of the 4th Corpus Linguistics Conference, University of Birmingham*, 27–30.
http://ucrel.lancs.ac.uk/publications/cl2007/paper/204_Paper.pdf
- Green, A., & Hawkey, R. (2012). Marking assessments: Rating scales and rubrics. In *The Cambridge Guide to Second Language Assessment* (pp. 299–306). Cambridge Univ. Press.
- Greeve, A. (2014). Summary of gamification. In *Proceedings of Moodle Moot Japan 2014* (Vol. 3, pp. 8–11).
http://moodlejapan.org/home/file.php/1/2014_Moot_files/MoodleMoot2014_Proceedings.pdf#page=8
- Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, 8(1), 23–34.
<https://doi.org/10.20982/tqmp.08.1.p023>

- Hinds, J. (1983). Contrastive rhetoric: Japanese and English. *Text - Interdisciplinary Journal for the Study of Discourse*, 3(2). <https://doi.org/10.1515/text.1.1983.3.2.183>
- Hirose, K. (2003). Comparing L1 and L2 organizational patterns in the argumentative writing of Japanese EFL students. *Journal of Second Language Writing*, 12(2), 181–209. [https://doi.org/10.1016/S1060-3743\(03\)00015-8](https://doi.org/10.1016/S1060-3743(03)00015-8)
- Hyland, K. (2003). *Second Language Writing*. Cambridge University Press.
- IELTS Writing tips for teachers—Grammatical range and accuracy | Take IELTS*. (n.d.). Retrieved April 21, 2020, from <https://takeielts.britishcouncil.org/teach-ielts/resources/videos/writing-grammar>
- IELTS Writing tips for teachers—Lexical resource | Take IELTS*. (n.d.). Retrieved April 21, 2020, from <https://takeielts.britishcouncil.org/teach-ielts/resources/videos/writing-lexical-resource>
- Kaplan, R. B. (1966). Cultural thought patterns in intercultural - cultural education. *Language Learning*, 16(1 - 2), 1–20.
- McKinley, J. (2013). Displaying Critical Thinking in EFL Academic Writing: A Discussion of Japanese to English Contrastive Rhetoric. *RELC Journal*, 44(2), 195–208. <https://doi.org/10.1177/0033688213488386>
- Parise, P. (2015). The three P's of increased student engagement in Moodle: An e-learning guide for teachers. *Bulletin of Kanagawa Prefectural Institute of Language and Culture Studies*, 4, 47–58.
- Parise, P. (2019). Adopting badges in Moodle to enhance engagement in e-Learning for in-service teachers of English. *Bulletin of Kanagawa Prefectural Institute of Language and Culture Studies*, 8, 31–42.
- Polio, C. (2012). How to Research Second Language Writing. In A. Mackey & S. M. Gass (Eds.), *Research Methods in Second Language Acquisition* (pp. 139–157). Wiley-Blackwell.
- Polio, C., & Friedman, D. A. (2017). *Understanding, evaluating, and conducting second language writing research*. Routledge.
- R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Reichelt, M. (2009). A critical evaluation of writing teaching programmes in different foreign language settings. *Writing in Foreign Language Contexts: Learning, Teaching, and Research*, 183–206.
- RStudio Team. (2019). *RStudio: Integrated Development Environment for R*. RStudio, Inc. <http://www.rstudio.com/>

- Sasajima, S., Nishino, T., Ehara, Y., & Nagamine, T. (2012). Aspects of Japanese EFL teachers' cognitions on communicative language teaching (CLT). *JACET 全国大会要綱*, 51, 146–147.
- Slomp, D. H. (2012). Challenges in assessing the development of writing ability: Theories, constructs and methods. *Assessing Writing*, 17(2), 81–91.
<https://doi.org/10.1016/j.asw.2012.02.001>
- Stapleton, P. (2002). Critical thinking in Japanese L2 writing: Rethinking tired constructs. *ELT Journal*, 56(3), 250–257. <https://doi.org/10.1093/elt/56.3.250>
- Stevens, D. D., & Levi, A. (2013). *Introduction to rubrics: An assessment tool to save grading time, convey effective feedback, and promote student learning* (2nd ed). Stylus.
- Tahira, M. (2012). Behind MEXT's new course of study guidelines. *The Language Teacher*, 36(3), 3–8.
- The IELTS Writing Rubric: What Does “Coherence and Cohesion” Mean? (2019a, August 26). *Graid*. <https://blog.graid.com/ielts-writing-coherence-and-cohesion/>
- The IELTS Writing Rubric: What Does “Task Achievement” Mean? (2019b, August 24). *Graid*. <https://blog.graid.com/task-achievement/>
- 文部科学省. (2018). 『【外国語編 英語編】高等学校学習指導要領(平成 30 年告示)解説』

APPENDIX A

The adapted rating scale based on the IELTS rubric

Rubric & Grading (Your grade will be highlighted in red.)

	Task Achievement	Coherence and Cohesion	Lexical Resource	Grammatical Range and Accuracy	Grade
5	Fully accomplishes all parts of the task and presents fully developed ideas.	Uses cohesion skillfully in such a way that attracts no attention.	Uses a wide range of vocabulary in an accurate and sophisticated manner.	Uses a wide range of structures with accuracy and flexibility.	
4	Addresses all parts of the task with clear ideas but could be more fully expanded.	Logically organizes information and ideas but may over/under-use some cohesive devices.	Uses a wide range of vocabulary fluently and flexibly to convey precise meanings. Makes occasional errors in word choice and collocation.	Uses a variety of complex structures, produces error-free sentences and has good control of grammar but may make a few errors.	
3	Paragraph addresses the task partially and expresses ideas which are not entirely clear nor well developed.	Presents information with some organization but makes errors in the use of cohesive devices.	Uses a limited range of vocabulary but this is minimally adequate for the task. Some errors cause difficulty for the reader.	Uses only a limited range of structures. Errors can cause some difficulty for the reader.	
2	Paragraph does not adequately address the task nor present any clear ideas.	Does not organize ideas logically.	Uses only a very limited range of words and expressions and errors severely distort the message.	Attempts to build sentences but errors in grammar and punctuation distort the meaning.	
1	Paragraph is completely unrelated to the task.	Fails to communicate a message.	Can only use a few isolated words.	Cannot make sentence forms at all.	
Draft					
Final					

<p>Grading Scale:</p> <p>16 – 20: A (Excellent)</p> <p>14 – 15: B (Very Good)</p> <p>12 – 13: C (Satisfactory)</p> <p>10 – 11: D (Marginal)</p> <p>0 – 9: F (Failure)</p>
--