

# Developing and Evaluating a Scoring Rubric for Argumentative Essays : A Module-Based Approach

|       |  |
|-------|--|
| メタデータ | 言語: English<br>出版者: Urban-Culture Research Center, Graduate School of Literature and Human Sciences, Osaka City University<br>公開日: 2021-07-09<br>キーワード (Ja): ルーブリック, モジュール, ライティング, 議論型エッセイ, 評価観点, 評価基準<br>キーワード (En): rubric, module, writing, argumentative essay, performance criteria, performance-level descriptions<br>作成者: 辻, 香代<br>メールアドレス:<br>所属: Osaka City University |
| URL   | <a href="https://ocu-omu.repo.nii.ac.jp/records/2017682">https://ocu-omu.repo.nii.ac.jp/records/2017682</a>  |

## Developing and Evaluating a Scoring Rubric for Argumentative Essays: A Module-Based Approach

Kayo TSUJI\*

Keywords: rubric, module, writing, argumentative essay, performance criteria, performance-level descriptions

### Explanatory Note

Rubrics have received increasing attention as assessment indicators to effectively evaluate students' complex performance and as an instructional tool to promote students' learning. Considering the positive effects of rubrics, the author attempts to develop an argumentative essay rubric. This study has two purposes: first, to design a scoring rubric incorporating students' writing challenges, and second, to evaluate the interrater reliability and validity of the rubric as an evaluation indicator. When using a rubric as a guide for improving text quality, it is necessary to set performance criteria and a clear task description as students need to know how and where to focus in an essay. As Fujishiro's (2009, 2011) *modular writing technique* is one effective approach to writing, the rubric should be developed based on this framework. The fundamentals of the modular approach support the decision of multiple criteria, that is, logic, clarity, and English. Logic emphasizes consistency between the respective modules, while clarity requires the inclusion of the necessary 5W1H information in each module. A third criterion, English, is relevant as the rubric applies to English-language texts. Each criterion includes the focus of evaluation and a six-level task description ranging from *ungradable* to *excellent*. Prior to educational use in the classroom, the validity and interrater reliability of the Module-based Writing Rubric were measured. Three evaluators rated 40 students' English written texts with the rubric and the Independent Writing Rubrics (Educational Testing Service, 2004). The validity of the rubric was measured for the 40 texts by comparing the value obtained by an evaluator for the Module-based Writing Rubric to the evaluator's value for the Independent Writing Rubrics. For the reliability test, the intraclass correlation coefficient for each performance criterion was calculated to confirm the degree of consistency among evaluators. The results confirm the validity and reliability of the newly developed rubric.

This article is a translated version of Tsuji's 2019 study. It was published in the *Journal of Japan Association for College and University Education*, 40 (2), when the author was affiliated with the Graduate School of Education, Kyoto University. The information on this practical report is as follows: Tsuji, K. (2019). Gironkata essei o hyōkasuru rūburikku no kōan to kentō: Mojūru o ishizue to shita raitingu gihō ni chakumokushite [Developing module-focused scoring rubrics for argumentative essays]. *Journal of Japan Association for College and University Education*, 40 (2), 64-71. Note that the English title and abstract of the present article have been partially revised to achieve better fluency.

### 1. Background and Purpose

In view of the globalized world of the 21st century, the Ministry of Education, Culture, Sports, Science and Technology in Japan (e.g. 2002, 2012) has placed strong emphasis on developing learners' English language skills. In particular, the country's Central Council for Education (2014) has highlighted the importance of cultivating the ability to communicate to a global audience in English. Of the four communication skills, writing may be particularly important. In the information age, Yamauchi (2010) states, writing is the everyday means for communicating across linguistic and cultural boundaries. Accordingly, this article focuses on an effective way of improving the writing skills required for an argumentative essay; the target task can contribute to cultivating learners' "ability to actively use English skills and assertively express [their] ideas (Central Council for Education, 2014).

While many factors constitute the foundation of English writing, the basis of writing, such as proper organiza-

\* Ph.D., Associate Professor, English Education Development Center, Osaka City University.  
E-mail: tsujikayo@osaka-cu.ac.jp

tion, is the most important for composing comprehensible texts. The same basis is required for compositions written in any language. More specifically, to achieve the purpose of written language, a writer must master what Mochizuki (2001, 2008) identified as the essential elements of writing: establishing convincing structure and content. Essays produced by student-writers, however, often fail to demonstrate these elements. As a means of clarifying for learners the fundamental requirements for the writing process, rubrics have garnered increasing attention (Nishitani, 2017).

Rubrics first emerged as a tool to assess learners' performance in the US in the 1980s. More than 20 years after their appearance in the US, researchers in Japan began studying rubrics. More recently, rubrics are being used not only to grade students' performance but also to support educational practice. Nishitani (2017), for instance, used a scoring rubric as a tool for training students' writing skills. Following this intervention, he found that learners gained a better understanding of the requirements for raising their writing quality, such as logical coherence and persuasiveness. Nishitani's 2017 study revealed a new horizon of possibilities for rubrics. In addition to providing evaluative criteria for scoring learners' comprehension and attainment, rubrics can also serve as a guide for improving learners' educational outcomes and, by indicating learning outcomes, provide a basis for learners to monitor their own achievements (Nishitani, 2017). Considering these opportunities, it is necessary that the learning requirements specified in a rubric be relevant to the particular writing challenges the learners face. Rather than using a generic rubric, a language educator needs to establish a unique rubric that considers learners' writing habits. Such a rubric tailored to the respective classroom potentially provides scaffolding to more effectively improve learners' performance on a writing task.

One set of rubrics relevant to the argumentative essay in English is the Independent Writing Rubrics (IWR) presented by the Educational Testing Service (ETS) in 2004. This set of rubrics provides a guide for scoring the writing ability of English as a Foreign Language (EFL) learners aspiring to study at an English-speaking university. While the IWR may apply to EFL learners in general, they are less relevant to the EFL milieu in Japan. Although they include task descriptions, they have no delineated performance criteria whereby Japanese EFL learners can tell what requirements they need to demonstrate. Wakita (2016) states that each writing course requires its own tailored rubric, one that reflects the learning objectives and expected outcomes for that course.

Rubrics offer clear benefits to teachers. By using metrics to measure students' achievement in specific writing tasks, teachers can monitor whether learners have mastered the learning requirements. They can then use this information to determine how to proceed with future teaching, including which aspects to focus on in subsequent lessons. Rubrics also benefit learners. Learners can use a rubric to gauge their level of attainment by identifying any aspect that they are still unsure of or the extent to which they have achieved the requirements. Tracking their progress in the task helps them determine how to proceed with future learning. In light of the abovementioned advantages of rubrics, an approach that tailors an assessment to students' particular writing challenges should have educational benefits.

A proficient argumentative essay should be incontrovertibly recognized as such. On this basis, when essays are graded by multiple assessors using the unique rubric designed in this study, each assessor should score the essay in a manner consistent with the prominent scoring guides used in global settings. Accordingly, the following premises are adopted for this study. First, language educators should identify the specific learning requirements their students need to master and then design bespoke rubrics that incorporate these requirements as learning objectives. Second, educators should assess whether their rubrics are effective in measuring performance and whether the values measured by the rubric are reproducible. Based on these premises, this study is intended to accomplish the following two objectives. The first objective is to design an original rubric reflecting the required learning outcomes. This rubric will define performance criteria as learning requirements that Japanese EFL students need to fulfill. Each criterion will include its definition, a scale showing the performance levels within that criterion, and descriptions of the requirements for each level. The second objective is to confirm the reliability and validity of the rubric. This task involves testing whether the rubric scores are reproducible and concordant with a comparable scoring rubric. The rubric used for this comparison should be the IWR, which is widely accepted in written English education. As mentioned previously, the IWR is a guide for scoring the writing section of the TOEFL iBT Test, an important milestone for EFL learners pursuing further study at an English-speaking university. The test measures the candidate's ability

to write an essay expressing his or her ideas on a particular subject (Council on International Educational Exchange [CIEE] Japan, 2011). This assessment index mirrors the requirements of an argumentative essay, the target task in this study, in both form and substance; writers must assert their claims clearly and support them with reasoning and concrete evidence.

## 2. Identifying Required Learning Outcomes

In creating a rubric, desirable learning outcomes should be determined. This process involves identifying the particular challenges Japanese EFL learners face in writing argumentative essays and defines what the students need to achieve to overcome them.

### (1) Challenges identified in students' essays

To achieve effective written communication in English, as Mochizuki (2001, 2008) argues, writers must establish clear content with logical flow. However, this is a daunting task for EFL learners. The following example of an argumentative essay, shown in Table 1, is indicative of typical challenges Japanese EFL students commonly experience. The excerpts from this essay, including the title, are quoted verbatim. This is an argument on the question of whether Japan should lower the minimum voting age to 18 years old. In the analysis below, the student's argument is paraphrased because the focus here is on the flaws in the argumentation rather than the language issues.

**Table 1** A Student's Argumentative Essay (Title: The right to vote from 18 years old)

|  |
|--|
| <b>[Claim]</b><br>I agree to lower the voting rights at the age of 18.   |
| <b>[Reasoning]</b><br>By lowering the voting rights, the generation gap is small. As more young people vote, politicians become conscious of young people. And, Politicians make policies for young people. Young people support Japan in the future. If we can develop policies about education and employment for young people, Japan is revitalized.  |
| <b>[Evidence 1]</b><br>Why the voting rights were lowered at 18 years? From the report of the Japanese Ministry of Internal Affairs, we found the reason. It is to reduce the intergenerational disparity. Currently Japan is an aging society with a declining birthrate, so most of the voting rights are occupied by people over the age of 60. By lowering the voting rights, the proportion of young people's voting rights will increase, so the generational gap will be small. And, younger people's opinions will be reflected. |
| <b>[Evidence 2]</b><br>The result of the questionnaire is the basis. Do you think that Japan will be better by increasing policies for young people? Yes: 65.9%. No: 34.1%. We set the criterion of the questionnaire result at 60% or more, so from this it can be said that Japan will be revitalized by increasing policies for young people.   |
| <b>[Conclusion]</b><br>Combining Basis 1 and basis 2 will lead to my reason of argument. Therefore, my argument is correct. So, I agree to lower the voting rights to the age of 18.   |

Examining the essay reveals that the student has failed to grasp the basic components of an argument; hence, parts of the essay are unconnected. The student asserts that the voting age should be lowered to 18. The reasoning is presented as follows: *Lowering the voting age reduces intergenerational inequality in voting enfranchisement. With a younger electorate, politicians will take young people more seriously and focus more on policies that benefit young people's lives. Youth represents the future of Japan, so if the government does more to improve youth education and career support, the country will have a brighter future.* For the two "evidence" fields, writers should present evi-

dence to justify the claim and reasoning. The student needed to cite cases, such as anecdotes, precedents, or raw data, illustrating how the public would benefit from less intergenerational inequality in voting enfranchisement. However, the response the student gave in the first evidence field, rather than providing such evidence, simply restated the reasoning: *Because Japan has an aging population, much of the electorate is over 60. Lowering the minimum voting age to 18 reduces intergenerational inequality and helps ensure that young people's views are incorporated into government policy.* Likewise, the answer given in the second evidence field fails the criteria for evidence: *In an opinion survey, more than 60% of young people said that Japan would have a brighter future if the government did more for young people.* The given statement offers no information on what educational or career policies would help rejuvenate Japan or how they would do so. For the conclusion, the student states the following: *My claim is correct. Therefore, I am in favor of lowering the voting age to 18 years old.* The offered evidence does not support the claim and reasoning, resulting in the conclusion lacking persuasiveness. Such illogical structures pervade Japanese EFL students' writing. In addition, as the descriptions of the evidence show, essential information concerning *who does what to whom, where, when, and how this is done* is not properly explained. The student shows little awareness of how to describe information for each component of their essay in a clear manner. This tendency has been attributed to the wider sociocultural environment. In short, Japan's ethnic and cultural homogeneity encourages Japanese EFL students to rely heavily on implicit understanding during communication (Nishimura *et al.*, 2008). This environment affects students' attitudes toward written communication, which leads to tenuous intelligibility in essays. Thus, the two key challenges above, identified from examining the student essay, demonstrate the areas in which Japanese ESL learners typically fail and that must be addressed to develop the learners' ability to effectively convey their ideas in writing.

## (2) Focus on a modular approach to writing

To overcome the challenges identified in the previous section, students must first grasp the interrelationships among the claim, reasoning, evidence, and conclusion. They must also understand the details to be described in each component. Stated differently, language educators should promote students' understanding of how their claim should be stated, what should be described as their reasoning behind it, what evidence should be adduced to support their claim and reasoning, and what conclusions should be drawn. One effective approach for encouraging the thinking process is a modular approach to writing, called the *modular writing technique* (Fujishiro, 2009, 2011). A module is an interchangeable component. Its etymology is from the Latin word *modulus*, which carries the now-archaic meaning of "the smallest unit" (Tsuji, 2018). As used in this article, modules are regarded as the smallest unit of an essay, that is to say, claim, reasoning, evidence, and conclusion. Fujishiro's modular approach, developed collaboratively between journalists and educators experienced in journalism, is effective for writing compelling articles (Fujishiro, 2009, 2011).

The approach is fundamentally designed to promote well-developed, well-elaborated writing that gets the intended message across to the audience. To this end, its basis involves the development of two techniques: 1) the technique to organize the structure of a composition such that one's audience can easily understand the overall flow of ideas and 2) the technique to elaborate on the points as necessary to ensure that the reader obtains a clear picture (Fujishiro, 2009, 2011; Namba, 2008). For the latter, the researchers emphasize the importance of following the five Ws and one H (5W1H: who does what, where, when, why, and how) for clear descriptions. In this study, this questioning technique should be considered as involving any interrogative element, not only the basic 5W1H elements of information. The *modularization* of writing allows students to examine how the modules connect with each other, resulting in greater overall consistency (Odanaka, 2013, p. 64). To relate this modular approach to the context of this study, the focus of the learning requirements should be twofold. The first requirement should focus on ensuring logical connections between the modules: claim, reasoning, evidence, and conclusion. More specifically, it should focus on identifying the most basic components necessary for a cohesive argument and discerning their logical relationships. The second requirement should focus on ensuring that the sentences within the modules are described for the respective modules to fulfill their functions. These two foci represent the required learning outcomes (Parts of this paragraph are taken from Tsuji [2018]).

### 3. Developing a Module-Based Writing Rubric

A modular approach should be applied to create the rubric. To determine the performance criteria and performance-level descriptions of the rubric, Ono and Matsushita's (2016) tabular writing rubric was adopted as a model. In this model, the required learning outcomes are incorporated as performance criteria. Each column represents a different performance criterion, and each row has a different performance level linked to a score. In addition, each performance level contains descriptions of what the student needs to demonstrate to attain that level. The rubric shown in Table 2 is a modified version of a writing rubric previously developed by Tsuji (2018). It will be referred to in the rest of this article as the Module-based Writing Rubric (MWR).

**Table 2** Module-based Writing Rubric (MWR)

|   | Logic: Establishing consistent logic through the writing   | Clarity: Achieving clarity in the writing  | English: Displaying competent language use and expression through the writing  |
|---|--|--|--|
| 5 | <p>A text at this level accomplishes the following:</p> <p>It effectively describes the arguments with a logical flow. It effectively states the claims related to the topic, and they are clearly supported with appropriate reasons and evidence. It effectively presents the conclusion that can be drawn from the arguments. Each module is relevant to the topic, and there is logical consistency between the modules.</p>                       | <p>A text at this level accomplishes the following:</p> <p>It effectively allows readers to develop a clear, concrete image of the information. It effectively describes/elaborates the arguments of each module with the clearly appropriate 5W1H information, and it effectively explains the unfamiliar words.</p>      | <p>A text at this level accomplishes the following:</p> <p>It effectively conveys the meaning behind a sentence and makes it easily comprehensible. It effectively conveys an idea in one sentence and renders it easily comprehensible. The subject is appropriately stated, and the main verb semantically and grammatically aligns with the subject. The parts of the sentence follow the correct grammatical forms of a sentence with very few mistakes. It effectively uses heads and modifiers, and it has very little redundancy in a sentence.</p> |
| 4 | <p>A text at this level accomplishes the following:</p> <p>It generally describes the arguments with a logical flow. It generally states the claims related to the topic, and they are necessarily supported with reasons and evidence. It generally presents a conclusion that can be drawn from the arguments. Each module is generally relevant to the topic, and there is substantial logical consistency between the modules.</p>                 | <p>A text at this level accomplishes the following:</p> <p>It generally allows readers to develop a clear, concrete image of the information. It generally describes/elaborates the arguments of each module with enough 5W1H information, and it generally explains the unfamiliar words.</p>                             | <p>A text at this level accomplishes the following:</p> <p>It generally conveys the meaning behind a sentence and makes it comprehensible. It generally conveys an idea in one sentence and renders it comprehensible. The subject is stated, and the main verb generally aligns semantically and grammatically with the subject. The parts of the sentence follow the grammatical forms of a sentence with occasional minor mistakes. It generally uses heads and modifiers, and it is slightly redundant in a sentence.</p>                              |
| 3 | <p>A text at this level accomplishes the following:</p> <p>It describes the arguments with a somewhat logical flow. It states the claims related to the topic, and they are somewhat supported with reasons and evidence. It presents the conclusion that can be drawn from the arguments, though some points are occasionally illogical. Each module is relevant to the topic, though there are some logical inconsistencies between the modules.</p> | <p>A text at this level accomplishes the following:</p> <p>It allows readers to develop a clear image of the information, though it is occasionally obscured. It describes/elaborates the arguments of each module with 5W1H information, though some of them are obscured. It somewhat explains the unfamiliar words.</p> | <p>A text at this level accomplishes the following:</p> <p>It conveys the meaning behind a sentence, though some points are occasionally unclear. It conveys an idea in one sentence and renders it readable. The subject and the main verb are stated, though semantically and grammatically mistakes have been occasionally made. The parts of the sentence somewhat follow the grammatical forms of a sentence. It uses heads and modifiers to some degree, and it is occasionally redundant in a sentence.</p>   |



|   | Logic: Establishing consistent logic through the writing  | Clarity: Achieving clarity in the writing   | English: Displaying competent language use and expression through the writing   |
|---|---|---|---|
| 2 | <p>A text at this level reveals one or more of the following:</p> <p>It describes the arguments with a limited logical flow. It attempts to state the claims related to the topic and to demonstrate reasons supporting the claims, but the reasons are inappropriate or insufficient. It attempts to describe the evidence, but the information provided is inappropriate or insufficient. It attempts to present the conclusion, but it is largely inconsistent with the arguments. It attempts to establish logical consistency between the modules, but there is limited consistency.</p> | <p>A text at this level reveals one or more of the following:</p> <p>It attempts to allow readers to develop a clear image of the information, but it is largely limited. It attempts to describe/elaborate the arguments of each module with 5W1H information, but they are inappropriate or insufficient. It attempts to explain unfamiliar words, but it is inappropriately or insufficiently explained.</p> | <p>A text at this level reveals one or more of the following:</p> <p>It attempts to convey the meaning behind a sentence, but it makes little sense. The subject is insufficiently aligned semantically and grammatically with the verb. The parts of the sentence insufficiently follow the grammatical forms of a sentence. It attempts to use heads and modifiers, but it is inappropriately used. It attempts to avoid redundancy, but it is largely redundant in a sentence.</p> |
| 1 | <p>A text at this level is seriously flawed due to one or more of the following:</p> <p>It describes the arguments with no logical flow. It fails to state the claims related to the topic and to demonstrate reasons supporting the claims. It fails to describe the evidence with little or no information, and it fails to present the conclusion. It establishes little or no logical consistency between the modules.</p>  | <p>A text at this level is seriously flawed due to one or more of the following:</p> <p>It fails to allow readers to develop an image of the information. It fails to describe/elaborate the arguments of each module with 5W1H information, and it fails to explain the unfamiliar words.</p>  | <p>A text at this level is seriously flawed due to one or more of the following:</p> <p>It fails to convey the meaning behind a sentence, and it makes no sense. The subject is not semantically or grammatically aligned with the verb. The parts of the sentence do not follow the necessary grammatical forms of a sentence. It fails to use heads and modifiers, and it fails to avoid redundancy in a sentence.</p>  |
| 0 | <p>A text at this level has little or no description. It has scant information for judgment.</p>  | <p>A text at this level has little or no description. It has scant information for judgment.</p>  | <p>A text at this level has little or no description. It has scant information for judgment.</p>  |

Note: The 5W1H includes all interrogative elements required for a clear argument.

## (1) Performance criteria and definitions

The first step was to incorporate the required learning outcomes into the rubric as performance criteria. The requirements were identified as follows. First, the writing achieves inter-module coherence, and second, it provides appropriate and sufficiently elaborated information for the modules to fulfill their function. On the basis that proper inter-module coherence is critical to establishing consistent logic, the first learning outcome, logical coherence, was included in the MWR as the performance criterion. It is expressed in the rubric as “Logic: Establishing consistent logic through the writing.” This criterion also covers the intra-module coherence. The reason for this is that intra-module coherence is essential for ensuring inter-module coherence. That is, insofar as inter-module coherence means connecting the modules of the essay, it requires students to compose the inner structure of each module in such a way as to ensure a logical progression of ideas from one module to the next. As the second learning outcome serves to embody each module’s description, this was incorporated into the MWR as clarity. It is expressed as “Clarity: Achieving clarity in the writing.” Alongside “Logic” and “Clarity,” the MWR added a third criterion in the rightmost column: “English: Displaying competent language use and expression through the writing.” This criterion is relevant as the rubric applies to English-language texts.

The next step was to define each performance criterion in the MWR. The renowned scoring rubric, the IWR (Educational Testing Service [ETS], 2004), consists of a holistic set of rubrics with only task descriptions for each performance level. With the holistic evaluation table, raters score based on the overall impression of their learner’s performance (Wakita, 2016, p. 24). More specifically, while the IWR presents an all-inclusive “task description” for the six levels, it neither delineates different performance criteria nor specifies the definition for any criterion. Accordingly, for the MWR’s first two performance criteria (“Logic” and “Clarity”), the definition of each criterion was

derived from previous research (e.g. Iguchi, 2008; Odanaka, 2013; Ono & Matsushita, 2016) specifying requirements for effective writing.

Regarding the definition of “Logic,” one source was Ono and Matsushita’s (2016) rubric, which was designed to assess Japanese-language academic writing. It uses six performance criteria. Of these, three resemble the criteria in the MWR: “Claim and Conclusion,” “Warrants and Facts/Data,” and “Overall Structure.” The first criterion is defined as “drawing a conclusion while associating the author’s assertion developed with respect to the issues of concern or interest.” The second criterion is defined as “providing warrants in support of the assertion and clarifying factual evidence demonstrating the veracity of the warrants.” The third criterion is defined as “ensuring logical coherence across the entirety of the writing, from raising a concerning theme to arriving at a conclusion, and expressing these ideas clearly” (Ono & Matsushita, 2016, pp. 32-33). These three criteria can be consolidated as follows: a proficient argumentative essay achieves the requirements of asserting a claim clearly, stating the reasoning for the claim, citing evidence in support of the claim and reasoning, and arriving at an appropriate conclusion. Another source for this criterion was Iguchi (2008): “A high-quality opinion essay has a clear claim and clear reasoning, has a logical structure, and contains persuasive assertions and logically consistent content” (p. 211). In essence, it indicates that an effective essay is logically constructed and clearly expressed from claim to conclusion. Including these definitions in the “modularized” (Odanaka, 2013, p. 64) argumentative essay, the result would be as follows: “The claim and reasoning are connected,” “the claim and reasoning are in accord with the evidence,” “the claim is consistent with the conclusion,” and “each component (module) is well described.” Thus, in light of the above research, the following definition was adopted for the MWR’s “Logic” criterion: *The claim is in accord with the reasoning; the reasoning is in accord with the evidence; the claim and reasoning are in accord with the evidence; the claim is consistent with the conclusion.*

The definition of “Clarity” was already supplied to some extent in part of Iguchi’s (2008) explanation, which can be translated into the modularized essay in which the arguments of each module are well elaborated. In addition to Iguchi, two other sources were consulted. The first was Odanaka (2013), who argued that writing is clearer if it includes more concrete information. The second was Regoniel (2016), who stated that good, clear writing provides answers to any questions readers are likely to have. Regoniel’s view resembles that of Fujishiro (2009, 2011) inasmuch as it underscores the importance of providing concrete details pertaining to basic 5W1H questions. To summarize these insights, for writing to achieve good clarity, each module (i.e. the claim, reasoning, evidence, and conclusion) must be reasonably well described and elaborated. The text must keep abstract information including lexical obscureness to a minimum by supplying the appropriate details. Thus, the following definition was adopted for the MWR’s “Clarity” criterion: *Each module is properly described, with clear explanations of any unfamiliar nomenclature. The text contains appropriate and clear 5W1H information. The text keeps abstract information to a minimum.* (Parts of this paragraph are taken from Tsuji [2018]).

## (2) Performance levels and descriptions

With the performance criteria defined, the next task was to determine the performance levels and their descriptions. As the MWR was designed to promote the ability to communicate globally, consistency with the IWR (ETS, 2004) scale was considered necessary. As mentioned, the IWR is an internationally recognized standard. Regarding the performance levels themselves, as shown in Figure 1, the IWR scores performance on a six-point scale (Score 0 to 5). The MWR adopted the same six-point scale as the IWR. Performance-level descriptions were added to the MWR in a manner that reflected the “task descriptions” of the six levels in the IWR. These task descriptions are outlined in Figure 1.

Starting with Score 5, the IWR reserves this top score for exemplary essays. The task description for this level describes the qualities of an excellent essay. Specifically, it requires that among other things, the essays be well organized, developed, and coherent, though they may have minor lexical or grammatical errors. The task descriptions for the other levels are summarized as follows: Essays scoring 4 are generally well organized and developed, although some minor points may not be fully elaborated and some trivial illogical parts can be seen. Essays scoring 3 often address the topic and required task, though the connection of ideas and some described information may be occa-



| Score | Task Description  |
|-------|---|
| 5     | <b>An essay at this level largely accomplishes all of the following:</b> <ul style="list-style-type: none"> <li>effectively addresses the topic and task</li> <li>is well organized and well developed, using clearly appropriate explanations, exemplifications, and/or details</li> <li>displays unity, progression, and coherence</li> <li>displays consistent facility in the use of language, demonstrating syntactic variety, appropriate word choice, and idiomaticity, though it may have minor lexical or grammatical errors</li> </ul>  |
| 4     | <b>An essay at this level largely accomplishes all of the following:</b> <ul style="list-style-type: none"> <li>addresses the topic and task well, though some points may not be fully elaborated</li> <li>is generally well organized and well developed, using appropriate and sufficient explanations, exemplifications, and/or details</li> <li>displays unity, progression, and coherence, though it may contain occasional redundancy, digression, or unclear connections</li> <li>displays facility in the use of language, demonstrating syntactic variety and range of vocabulary, though it will probably have occasional noticeable minor errors in structure, word form, or use of idiomatic language that do not interfere with meaning</li> </ul> |
| 3     | <b>An essay at this level is marked by one or more of the following:</b> <ul style="list-style-type: none"> <li>addresses the topic and task using somewhat developed explanations, exemplifications, and/or details</li> <li>displays unity, progression, and coherence, though connection of ideas may be occasionally obscured</li> <li>may demonstrate inconsistent facility in sentence formation and word choice that may result in lack of clarity and occasionally obscure meaning</li> <li>may display accurate but limited range of syntactic structures and vocabulary</li> </ul>  |
| 2     | <b>An essay at this level may reveal one or more of the following weaknesses:</b> <ul style="list-style-type: none"> <li>limited development in response to the topic and task</li> <li>inadequate organization or connection of ideas</li> <li>inappropriate or insufficient exemplifications, explanations, or details to support or illustrate generalizations in response to the task</li> <li>a noticeably inappropriate choice of words or word forms</li> <li>an accumulation of errors in sentence structure and/or usage</li> </ul>  |
| 1     | <b>An essay at this level is seriously flawed by one or more of the following weaknesses:</b> <ul style="list-style-type: none"> <li>serious disorganization or underdevelopment</li> <li>little or no detail, or irrelevant specifics, or questionable responsiveness to the task</li> <li>serious and frequent errors in sentence structure or usage</li> </ul>   |
| 0     | <b>An essay at this level</b> merely copies words from the topic, rejects the topic, or is otherwise not connected to the topic, is written in a foreign language, consists of keystroke characters, or is blank.   |

**Figure 1** Independent Writing Rubrics (ETS, 2004)

sionally obscured. Essays scoring 2 display limited development in response to the topic and task, inadequate organization, and insufficient or irrelevant supporting explanations. Essays at this level require significant improvement. Essays scoring 1 display serious disorganization or underdevelopment, indicating severely deficient writing. As for the bottommost rung (essays with a score of 0), this is reserved for submissions that fail the most basic requirements of an essay. It is a subscale level used when writing is determined to be of extremely poor quality. Examples include answers that are largely blank.

Based on the IWR scale, the MWR adopted the same six levels for its three performance criteria, labeling the scales as 0 (*ungradable*), 1 (*deficient*), 2 (*improvement required*), 3 (*acceptable*), 4 (*good*), and 5 (*excellent*). Each performance-level description for each criterion of the MWR was determined to correspond to the task description of each level in the IWR.

For the third criterion, the MWR derived performance-level descriptions from those in the IWR related to language, syntax, and grammar. The IWR's task description for Score 5 states that an essay at this level "displays consistent facility in the use of language, demonstrating syntactic variety, appropriate word choice, and idiomaticity, though it may have minor lexical or grammatical errors." For the MWR, descriptions concerning the basic structure of English sentences were adopted in view of the challenges specific to Japanese EFL learners. The descriptions for each level in the "English" criterion were determined, like those for the other two performance criteria, to align with the IWR's task descriptions. However, it should be noted that the descriptions do not include cases in which errors are automatically corrected because the students participating in this study produced their essays on word-processing software. In addition, this article is more concerned with the first two language-spanning criteria; a detailed explanation regarding the definition of the English criterion is omitted here.

Using the above procedure, the MWR was developed with three performance criteria (Logic, Clarity, and English), each rated on a six-point scale (Score 0 to 5: *ungradable* to *excellent*). A score of 5 (*excellent*) on all three criteria would result in the best total score of 15. Essays scored at a score of 0 (*ungradable*) on all three criteria would end up with a total score of 0. The rubric also allows evaluators to score an essay between two levels. For example, the evaluator can give a score of 2.5 for a case where the essay is not at a passing level but is higher than a score of 2 (*improvement required*). The above evaluation was used in the MWR because the IWR also allows for scoring between two levels.

#### 4. Validity and Reliability Testing Methodology

Having developed the MWR, the subsequent phase was to test its validity and reliability. In terms of the validity, the content- and criterion-related validity were tested by three evaluators (Evaluators A, B, and C), one of whom (Evaluator A) was the creator of the scale. For content validity, the definitions of the performance criteria were compared to those used in the IWR. For criterion-related validity, the measurement adopted the method in Takahashi (2012), in which results from the model of interest are compared to those on a comparative model to determine their correlation (p. 219). Each of the three evaluators scored 40 English-written essays using both the MWR and IWR. Then, the MWR scores of each evaluator were compared to that evaluator's IWR scores in order to check the correlation for each individual evaluator.

Reliability was measured based on interrater reproducibility, that is, the extent to which the scores remain consistent and stable between different evaluators (Takahashi, 2012, p. 217). For this approach, it is necessary to determine whether scores given by different evaluators are consistent (Eguchi, 2012, p. 82). Accordingly, after the three evaluators had scored the 40 English-written texts using the MWR, the intraclass correlation coefficient (ICC) for each performance criterion was calculated to verify the degree of consistency among evaluators. The MWR's first two criteria, Logic and Clarity, were additionally tested on Japanese-written argumentative essays on the basis that these criteria describe the fundamentals required for proficient essays written in any language. Namely, two evaluators (Evaluators A and D: native Japanese) scored 50 Japanese essays using the MWR, and the two sets of scores were compared to verify the reliability of the logic and clarity criteria.

#### 5. Results of the Validity Assessment

##### (1) Content validity

To assess content validity, an excellent argumentative essay was first conceptualized based on the task description for awarding an exemplary essay (Score 5) under the IWR. Next, its content was carefully scrutinized to evaluate the consistency with the performance-level description for an excellent essay (Score 5) under the MWR. Part of the IWR's task description for Score 5 indicates that an essay at this level is "well-organized and well-developed, using clearly appropriate explanations, exemplifications, and/or details." This description was translated into the modular framework to *provide appropriate evidence (i.e. explanations and exemplifications) to support the claim and reasoning and to describe appropriate details by clarifying the necessary 5W1H information*. In light of this

modularized definition, the IWR was deemed to encompass the logic and clarity criteria of the MWR. Furthermore, the IWR's task description for Score 5 indicates that an essay at this level "displays unity, progression, and coherence." This can be judged to correspond to the first criterion of the MWR, "Logic: Establishing a consistent logic through the writing." Stated differently, the IWR's description was, in turn, interpreted as *displaying smooth logical progression from claim to reasoning, to evidence, and finally to the conclusion*. The three evaluators agreed with this interpretation and thus concurred that the MWR has content validity.

## (2) Criterion-related validity

The validity of the MWR was evaluated by measuring its correlation with the IWR for each individual evaluator. The results of the correlation analysis between the scores measured by both rubrics are described as follows: the coefficient for Evaluator A's two sets of scores was 0.86; for Evaluator B, it was 0.90; and for Evaluator C, it was 0.94. Simply put, regardless of the evaluator, the two sets of scores exhibited a strong positive correlation. These findings indicate that the MWR has criterion-related validity.

## 6. Results of the Reliability Assessment

Ono and Matsushita (2016) highlighted the issue of evaluator subjectivity in scoring, arguing that steps should be taken to maximize inter-rater reliability. Accordingly, before scoring the essays, the three evaluators jointly examined the content validity of the MWR and established a shared understanding of its performance criteria, including the definitions thereof, the six-point scoring scale, and performance-level descriptions for each criterion.

The ICCs for the three evaluators' scores were 0.77 for Logic, 0.69 for Clarity, 0.66 for English, and 0.77 for the three criteria combined. The 95% confidence intervals for each were as follows: for Logic, the lower limit was 0.65 and the upper limit was 0.86; for Clarity, they were 0.54 and 0.81; for English, they were 0.51 and 0.79; and for the three criteria combined, they were 0.65 and 0.86. Cohen's kappa coefficient and other measures of reliability define a reliable model as one that achieves an ICC of 0.6 or higher and a highly reliable model as one that achieves an ICC of 0.7 or more (Tsushima, 2016). On this basis, the MWR can be considered to demonstrate inter-rater reproducibility.

As part of the reliability test for the first two criteria, the two evaluators (Evaluators A and D) additionally assessed 50 Japanese written essays. As stated previously, these criteria describe language-spanning requirements. The two sets of scores showed strong positive correlations. For Logic, the correlation between the two evaluators' scores was 0.82; for Clarity, it was 0.81; and for the two criteria combined, it was 0.83. These results indicate that the MWR's measurement of the two criteria remains consistent across languages; therefore, the rubric may offer a reliable measure even for assessing Japanese written argumentative essays.

The ICC results for Clarity (0.69) and English (0.66), while not markedly low, fell just short of the threshold for a highly reliable ICC value, which is 0.70. The latter criterion affects the former, as the ability to compose English sentences is strongly related to the ability to compose clear sentences in that language. The IWR's task description for Score 3 includes the following description (italicization by author): An essay at this level "may demonstrate inconsistent facility in sentence formation and word choice that *may result in lack of clarity and occasionally obscure meaning*." Such sentence formation and word choice are illustrated in the following student's sentence, which each evaluator interpreted and evaluated differently: "The books are controlled as data. So, we can easily search for objective books." Although the sentences contain several unnatural expressions and grammatical errors, Evaluator A understood the writer's intended message and interpreted it to mean, "*Since the books have been digitized, it is easier to find the book we want.*" However, Evaluators B and C decided that the first part of the sentence, "The books are controlled as data," was unclear and thus felt that the writer's intention was completely obscure. This was due in part to the grammatically inconsistent sentence encompassing an unnatural vocabulary choice that followed this part. There were other cases of such inter-evaluator discrepancy, and the culmination of such cases likely impacted the ICCs for both Clarity and English. For English, which had the lowest ICCs of the three performance criteria, the average measured values were 2.65, 2.43, and 2.45 for evaluators A, B, and C, respectively. The mean values



for Clarity were 2.53, 2.35, and 2.46. These results demonstrate that the measured values from Evaluator A were relatively higher with respect to both criteria. The degree of the evaluators' understanding of the students' work may affect the outcome of such evaluations. Evaluator A is a native speaker of Japanese, born and raised in Japan, who is well acquainted with the circumstances particular to Japanese EFL learners. As such, this evaluator keenly understands the broader cultural and historical context behind the subject matter. Moreover, Evaluator A recognizes the features of English writing produced by Japanese learners. In contrast, Evaluators B and C are non-native Japanese speakers who moved to Japan after spending several years in the West following graduation. Controlling for such differences in evaluator background and the accompanying subjectivity is critical to ensure neutrality. A variety of appropriate procedures should be considered to achieve objectivity in future research and educational settings.

## 7. Conclusion

This article began from the premise that because Japanese EFL learners poorly grasp the most basic elements of cohesive writing, it is necessary to provide a writing guide emphasizing the interrelationships among these elements (i.e. claim, reasoning, evidence, and conclusion) as well as the ways in which to describe each element. On this basis, a rubric founded on Fujishiro's (2009, 2011) modular writing technique was formulated. The resulting rubric, the Module-based Writing Rubric, consists of three performance criteria: "Logic: Establishing consistent logic through the writing," "Clarity: Achieving clarity in the writing," and "English: Displaying competent language use and expression through the writing." The validity and reliability of the MWR were examined, revealing that the rubric's criteria incorporating the particular learning challenges, the operationalized definitions of these criteria, its six-point scale, and the performance-level descriptions for each criterion may all be deemed appropriate. It is hoped that this initiative serves as an insightful example of a writing index reflecting the learning outcomes required for EFL learners.

## Acknowledgements

I would like to offer my sincerest thanks to Associate Professor Tsuyoshi YAMADA of Kyoto University (currently a professor at Kansai University) for his valuable guidance.

## References

- Central Council for Education. (2014). *Atarashii jidai ni fusawashii kōdai setsuzoku no jitsugen ni muketa kōtōgakkō kyōiku, daigaku kyōiku, daigaku nyūgakusha senbatsu no ittaiteki kaikaku ni tsuite: Subete no wakamono ga yume ya mokuhyō o mebukase, mirai ni hanahirakaseru tame ni (tōshin)* [On integrated reforms in high school and university education and university entrance examinations aimed at realizing a high school and university articulation system appropriate for a new era: Creating a future for the realization of the dreams and goals of all young people]. Central Education Council's Report. [http://www.mext.go.jp/b\\_menu/shingi/chukyo/chukyo0/toushin/icsFiles/afieldfile/2015/01/14/1354191.pdf](http://www.mext.go.jp/b_menu/shingi/chukyo/chukyo0/toushin/icsFiles/afieldfile/2015/01/14/1354191.pdf)
- Council on International Educational Exchange, Japan. (2011). TOEFL® tesuto gaiyō [Overview of the TOEFL test]. <https://www.cieej.or.jp/toefl/toefl/format.html>
- Educational Testing Service. (2004). *IBT/Next Generation TOEFL Test: Independent Writing Rubrics (scoring standards)*. [http://www.ets.org/Media/Tests/TOEFL/pdf/Writing\\_Rubrics.pdf](http://www.ets.org/Media/Tests/TOEFL/pdf/Writing_Rubrics.pdf)
- Eguchi, M. (2012). Naiyō bunseki no shuhō o mochiita raitingu rūburikku gutaika no kokoromi [Creating specific rubrics for writing English sentences using content analysis]. *Shimane Journal of Policy Studies*, 24, 73-84.
- Fujishiro, H. (2009). "Mojūru raitingu" no kaiatsu to bunshō kyōiku ni okeru jissen jirei [Development of "module writing" and its application to writing education]. *Unit for Science and Technology Communication, Hokkaido University*,

6, 92-101.

- Fujishiro, H. (2011). *Hasshinryoku no kitaekata sōsharu media katsuyōjutsu* [Honing communication skills: Ability to use social media]. PHP Institute.
- Iguchi, A. (2008). Chūgakusei no ikenbun sakusei katei ni okeru metaninchi hōryaku no kyōikuteki datōsei no kentō [Educational validity of metacognitive strategy used by junior high students when they write argumentative essays]. *Bulletin of the Graduate School of Education, Hiroshima University, Part 2: Arts and Science Education*, 57, 209-218.
- Ministry of Education, Culture, Sports, Science and Technology. (2002). “*Eigo ga tsukaeru Nihonjin*” no ikusei no tame no senryaku kōsō no sakutei nitsuite [Establishing a strategic vision for fostering English-proficient Japanese citizens]. [http://www.mext.go.jp/b\\_menu/shingi/chousa/shotou/020/sesaku](http://www.mext.go.jp/b_menu/shingi/chousa/shotou/020/sesaku)
- Mochizuki, A. (2001). *Shin gakushū shidō yōryō ni motozuku eigoka kyōiku hō* [An approach to English language teaching based on the new curriculum guidelines]. Taishukan Shoten.
- Mochizuki, A. (2008). Purosesu raitingu no yūkōsei: Daigakusei no baai [The efficacy of process writing: The case of university students]. *J Department of Education, Teikyo University*, 33, 37-47.
- Namba, M. (2008). Kōtō kyōiku ni okeru saiensu raitingu kyōiku: Sono hitsuyōsei to kōka [Science Writing Education in the Higher Education Institutions: Needs and Effect]. *Journal of Science Communication*, 4, 101-115.
- Nishimura, S., Nevgi, A., & Tella, S. (2008, February). Communication style and cultural features in high/low-context communication cultures: A case study of Finland, Japan, and India. In *Proceedings of a Subject-Didactic Symposium, Finland, Helsinki*.
- Nishitani, H. (2017). Bunshōryoku yōsei no tame no rūburikku katsuyō no kyōikuteki igi no kentō: Jugyō jissen kara miru kyōiku shuhō [Considering the educational significance of rubrics in academic writing: An educational method to judge from lecture practice]. *Kyoto University Journal of Higher Education*, 23, 25-35.
- Odanaka, A. (2013). Mojūru ni motozuita shōronbun sakusei gihō [Modular technique for essay writing]. In Kansai Faculty Development Association, Kyoto University Center for the Promotion of Excellence in Higher Education (Eds.) *Shikōshi hyōgensuru gakusei o sodateru raitingu shidō no hinto* [Insights for writing instruction to foster inquisitive and expressive students] (pp. 58-75). Minerva Shobo.
- Ono, K., & Matsushita, K. (2016). Shonenji kyōiku ni okeru repōto hyōka [Evaluating reports for first-year teaching]. In K. Matsushita, & T. Ishii (Eds.), *Akutibu rāningu no hyōka* [Active learning evaluation] (pp. 26-43). Toshindo.
- Project for Promotion of Global Human Resource Development. (2012). *Gurōbaru jinzai ikusei senryaku* [Strategy for fostering global individuals]. <http://www.kantei.go.jp/jp/singi/global/1206011matome.pdf>
- Regoniell, P. A. (2016). *How to write a thesis in the information age*. Metro Manila: Books on Demand Philippines, Inc.
- Takahashi, Y. (2012). Kyōiku sokutei hyōka bamen ni okeru shinraisei to datōsei [Reliability and validity in the context of measuring/assessing teaching]. In Center for the Promotion of Excellence in Higher Education, Kyoto University (Ed.), *Seiseisuru daigaku kyōikugaku* [Emerging university pedagogy] (pp. 217-222). Nakanishiya Shuppan.
- Tsuji, K. (2018). Gaikokugo raitingu ni okeru bogo bunshō keisei katsudō no kyōikuteki igi no kentō: Gakusei no gakushū ninshiki ni oyobosu eikyō ni shōten o atete [Examining the effects of first-language formulating activity during second-language writing on student perceived learning]. *Language Education & Technology*, 55, 247-276.
- Tsuji, K. (2019). Girongata essei o hyōkasuru rūburikku no kōan to kentō: Mojūru o ishizue to shita raitingu gihō ni chakumokushite [Developing module-focused scoring rubrics for argumentative essays]. *Journal of Japan Association for College and University Education*, 40(2), 64-71.
- Tsushima, E. (2016). *SPSS de manabu iryōkei deita kaiseki, dai 2 ban* [Using SPSS to learn medical data analysis, second edition]. Tokyo Toshō.



- Wakita, R. (2016). Raitingu rūburikku no jissen [Educational practice of the writing rubric]. *Doshisha Studies in Global Communications*, 5, 21-50.
- Yamauchi, Y. (2010). *Kaku chikara o sodateru daigaku kyōiku no paneru disukasshon* [Panel discussion on fostering writing skills in higher education]. <http://fukutake.iii.u-tokyo.ac.jp/archives/beat/seminar/044.html>