

**TEST REVIEW: THE VALIDITY OF WRITING SECTIONS ACROSS FIVE GRADES  
IN THE EIKEN TEST  
HIKARU ISHIYAMA**

*University of Hawai'i at Mānoa*

**INTRODUCTION**

Despite the widespread use of English standardized tests such as the *Test of English as a Foreign Language* (TOEFL) and the *International English Language Testing System* (IELTS), some countries in the so-called outer and expanding circles (Kachru, 1985) still have their own English language proficiency tests. The *Eiken Test in Practical English Proficiency* (実用英語技能検定 *jitsuyo eigo ginou kentei*; hereafter Eiken) was developed in 1963 by the *Society for Testing English Proficiency* (STEP), a Japanese non-profit organization affiliated with the Ministry of Education, Culture, Sports, Science, and Technology (MEXT). With the governmental support, the test has achieved prominence in Japan. The number of test takers has increased as the use of the test results has been expanded. During the 2020 school year, the number of test takers reached more than three million (more information available at <https://www.eiken.or.jp/eiken/merit/situation/>). The upper levels among the seven levels of Eiken called *Grade* have been used as a language proficiency test for university admissions not only for Japanese universities but also for the University of Hawaii and some other English-dominant universities overseas. Given that tests can give a considerable influence on what is taught at school (i.e., washback; Bailey, 1996), it is imperative to examine whether it measures what it is supposed to measure as well as if the test use is appropriate in the target context (Chapelle, 2012). Previous literature examining Eiken tests predominately focused on the reading and listening sections (Chujo & Oghigian, 2009; Hamada, 2015; Miura & Beglar, 2002; Piggin, 2011; Plumb & Watanabe, 2016). This may reflect the Japanese examination culture, where receptive skills (i.e., reading and listening) are prioritized over productive skills (i.e., speaking and writing). To fill the lacuna in the review of Eiken tests, the present paper focuses on the validity of writing sections across levels and provides information for stakeholders in terms of the use of the test results for their purposes.

## TEST PURPOSE AND USE

Eiken covers a range of proficiency levels, from elementary to advanced, divided into seven distinctive levels called *Grade*: Grade 5, 4, 3, Pre-2, 2, Pre-1, and 1. Each Grade measures English proficiency in four skills (i.e., listening, reading, speaking, and writing) except for the lowest two levels, which measure only listening and reading. STEP provides the correspondence of each Grade to school grades on its website (<https://www.eiken.or.jp/eiken/exam/about/>). As the targets of test takers are different depending on Grades, the use of each test varies from an achievement test for school graders to a proficiency test for college admissions. Because of the association between school curriculum and Eiken Grades, a large number of Japanese secondary schools have conventionally used Eiken as achievement tests up to Grade 2 (high school graduation level). According to the report provided by STEP (<https://www.eiken.or.jp/eiken/merit/situation/>), among all test takers in 2020, the number of “secondary school students” was the highest (2,911,389), followed by “others” (383,598), “elementary and below” (325,390), and “college students” (57,784).

Since STEP promoted Eiken use as an academic proficiency test in 2007 at the JALT conference, a number of tertiary schools have accepted the top three Grades as language proficiency certificates for university admissions (Piggin, 2011). For example, the applicants who have a certain Grade are given additional points to their scores of a college entrance exam or are even exempted from taking entrance exams. Not only Japanese universities but also universities overseas have recognized Eiken as a good indicator of applicants’ language proficiency levels. As of April 2021, over a hundred four-year universities in the U.S., as well as a large number of public high schools in Australia, have accepted the upper three Grades (i.e., Grade 2, Pre-1, and 1) (<https://www.eiken.or.jp/eiken/abroad/about/>). Likewise, since the University of Hawaii-West O’ahu accepted Eiken as a language assessment for international applicants in 2018, all campuses of the University of Hawaii have been using Eiken results as evidence of English proficiency. In addition to advantages for college admissions, the holders of the upper Grades benefit in career opportunities as well. For example, the Japan Tourism Agency supported by the Japanese government allows Grade 1 holders to be exempted from taking a language test (Japan National Tourism Organization, 2022).

## TEST METHOD

As Table 1 shows, each grade has a different test structure (STEP, n.d.). The number bolded indicates the number of items in each task and the time in parenthesis shows the time allotted for each task. For example, Grade Pre-2 has 37 items of vocabulary and reading comprehension in the reading section, one independent writing task in the writing section, 30 items in the listening section, and five items in the speaking section. Reading and writing are integrated as one section, so examinees need to complete both sections in 75 minutes. All questions in the reading and listening sections are multiple-choice and examinees mark one of four options on their answer sheet. As for the writing section, the examinees write with a pencil on the back of their mark sheet. According to information provided on the website, writing is graded by a trained rater with an analytical rating (STEP, n.d.). The examinees first receive the result of the first stage comprised of reading, writing, and listening. Only those who passed the first stage can take the speaking test about a month later at an appointed test center. The speaking test is conducted face-to-face with one examiner except for Grade 1, to which two examiners are assigned.

**Table 1**

*The Eiken Test Format (adapted from <https://www.eiken.or.jp/eiken/exam/>)*

	Reading + Writing	Listening	Speaking	Total
1	<b>41 + 1</b> Essay Writing (100 min)	<b>27</b> (35 min)	<b>1</b> (10 min)	145 min
Pre-1	<b>41 + 1</b> Essay Writing (90 min)	<b>29</b> (30 min)	<b>5</b> (8 min)	128 min
2	<b>38 + 1</b> Essay Writing (85 min)	<b>30</b> (25 min)	<b>5</b> (7 min)	117 min
Pre-2	<b>37 + 1</b> Sentence Writing (75 min)	<b>30</b> (25 min)	<b>6</b> (6 min)	106 min
3	<b>40 + 1</b> Sentence Writing (50 min)	<b>30</b> (25 min)	<b>6</b> (5 min)	80 min
4	<b>35</b> (35 min) No writing	<b>30</b> (30 min)	N/A	65 min
5	<b>25</b> (25 min) No writing	<b>25</b> (25 min)	N/A	50 min

The test is administered three times a year, January, May, and October. While student examinees normally take the test in their own classrooms, non-school-based examinees take the test at a test center. More than 400 test centers are placed across Japan. Moreover, due to the expanded use of the test for university admissions overseas, Eiken tests are now administered in London, New York, Los Angeles, and Honolulu.

## SCORES

As for the scoring system, STEP provides comprehensive information on its website (available at <https://www.eiken.or.jp/cse/>). The Eiken test is a pass-or-fail system based on the *Common Scale of English* (CSE), a scoring system originally developed by STEP in 2014. The CSE score sets 4,000 points at the highest language proficiency level and zero at the lowest. Grade Pre-1 holders, for example, are in the range of 2,305 points to 3,000 points. STEP also assigns a range of points to each of the bands in *Common European Framework of Reference for Languages* (CEFR). For example, C2 level is from 3,300 points to 4,000 points, C1 is from 2,600 points to 3,300 points, B2 is from 2,300 points to 2,600 points, and it goes on to A1. The association of the Eiken CSE score and the CEFR made different language tests comparable to each other. For instance, Grade Pre-1 holders have at least 2,305 points in the CSE score, which corresponds to B2 in CEFR. As CEFR bands correspond to other standardized tests, Eiken's results can be compared to scores from other standardized tests such as TOEFL and IELTS.

Since the implementation of CSE scores into Eiken, the result for a pass or a fail has been decided based on a cut-score. The cut score is calculated from tests administered previously and is different depending on the Grade but fixed at each Grade. For example, the cut-score for the first stage (i.e., reading, listening, and writing) of Grade 3 is 1103 points, and the one for the second stage (i.e., speaking) is 353 points. It means that the cut-score of Grade 3 is 1456 points in total out of 2200 points. Each of the four skills has the same full score so that the same value is weighed for each skill. Grade 1, for example, has 3,400 points in total and 850 points for each skill. The test takers receive their scores in each task as well as the pass-or-fail result.

## TEST COST

The cost is different depending on the Grade: 11,800 yen for Grade 1; 9,800 yen for Grade Pre-1; 8,400 yen for Grade 2; 7,900 yen for Grade Pre-2; 6,400 yen for Grade 3; 4,500 yen for Grade 4; 3,900 yen for Grade 5 (as of June 2022; <https://www.eiken.or.jp/eiken/schedule/>). It is relatively reasonable compared to other language tests. Even the highest cost for Grade 1 is less than half the price of TOEFL iBT. This explains one of the reasons Eiken is still popular among

Japanese people. More information about the test is available both in Japanese (<https://www.Eiken.or.jp/Eiken/>) and in English (<https://www.Eiken.or.jp/Eiken/en/>).

## VALIDITY

Against the backdrop of expanding the use of results from the Eiken examination, the validity of the Eiken test has been examined both by its developers (Brown et al., 2010) and researchers (Chujo & Oghigian, 2009; Hamada, 2015; Miura & Beglar, 2002; Piggini, 2011; Plumb & Watanabe, 2016). Most of these studies, however, focused on vocabulary, reading, and listening sections. No review has been conducted on the writing section of the Eiken examination to the extent of my knowledge. Moreover, regardless of the seven-staged examination, no discussion has been published to examine how the levels of each grade are structured and distinguished from each other.

To fill in the gap in the previous literature, the rest of the present paper will be devoted to considering the validity of the writing sections across levels. Validity in language testing can be broadly defined as “an evaluation of the credibility, or plausibility, of the proposed interpretations and uses of test scores” (Kane, 2010, p. 180). To establish sound validity, test makers should explicitly propose what the test measures as well as how the results of the test should be used for any decision making. In a book overviewing writing tasks in the Cambridge suite of examinations, Shaw and Weir (2007) proposed five key factors for the writing test validation. They include *cognitive validity*, *context validity*, *criterion-related validity*, *scoring validity*, and *consequential validity*. In this paper, *cognitive validity* and *criterion-related validity* will be excluded from the focus as both need the examination of the examinees’ actual writing performance. *Cognitive validity* is defined as “how closely [a writing task] represents the cognitive processing involved in writing contexts beyond the test itself” (Shaw & Weir, 2007, p. 34). To validate it, researchers need to investigate how examinees employ cognitive processing such as macro-planning, organization, micro-planning, translation, monitoring, and revising while they perform a writing task (Guapacha Chamorro, 2022; Shaw & Weir, 2007). *Criterion-related validity* is concerned with “the extent to which test scores correlate with a suitable external criterion of performance with established properties” (Shaw & Weir, 2007, p. 6). Criterion-related validity also requires the examinees’ performance. Thus, the present paper will

only review the other three aspects of validity (i.e., context validity, scoring validity, and consequential validity).

### ***Context Validity***

*Context validity* is the term employed by Shaw and Weir (2007), referring to the traditional notion of *content validity*. In their definition, *context validity* “relates to the linguistic and content demands that must be met for successful task realization and to features of the task setting that serve to describe the performance required” (Shaw & Weir, 2007, p. 63). Linguistic demands refer to test takers’ ability to demonstrate lexical and structural resources, discourse mode, and content knowledge. Task settings refer to the conditions such as instructions, writing purpose, text length, time allotment, writer-reader relationship, and physical conditions (Shaw & Weir, 2007). Each grade in Eiken is created in accordance with a criterion of writing skill (STEP, n.d.), as is shown in Table 2. The upper two levels, for example, have the terms *logically* in the statement. This indicates that the examinees are expected to write an essay with a certain amount of words with coherence and cohesion. Although each writing construct has different words, it is hard to differentiate from one another. For example, the only difference in the constructs between Grade 1 and Grade Pre 1 is the word “*wide*.” This may not be sufficient to proclaim that Grade 1 holders have different levels of writing ability from Grade-Pre 1 holders. STEP should provide more concrete writing constructs for the different five grades. The discussion below is going to address how each criterion is operationalized in writing tasks of each grade.

**Table 2**

*Criteria of Writing Skills of Eiken Grades (adapted from*

*<https://www.eiken.or.jp/eiken/exam/criteria/>)*

1	Can write logically on a wide range of socially relevant topics
Pre-1	Can write logically on a range of socially relevant topics
2	Can write on socially relevant topics
Pre-2	Can write on topics from everyday life
3	Can write simple texts about himself/herself

**Linguistic Demands.** Five levels of writing tasks are individualized in the writing prompts. The prompts shown in Table 3 were taken from the previous Eiken tests held in January 2021 (Previous Eiken tests are available on its website; <https://www.eiken.or.jp/eiken/exam/>). The main difference between the upper two levels and the lower three levels is the expected writing genre. Examinees of Grade 1 and Grade Pre-1 are specifically instructed to write an argumentative essay with a traditional essay structure: introduction, main body, and conclusion. Given that most secondary school students rarely have a chance to learn an essay structure either in their first language or second language (Kobayashi & Rinnert, 2002), it is reasonable that only the top two grades require a clear essay structure. For the lower three levels, examinees are only expected to provide their opinion with two reasons to support it.

The level of vocabulary used in prompts and the *topic* of each grade also differentiate the upper grades from the lower. The *topic* for the Grade 1 examination held in January 2021, for example, was, “Are economic sanctions a useful foreign-policy tool?” If examinees do not know the meaning of *sanctions*, they may not be able to write anything. In addition to broad lexical knowledge, they also need to have content knowledge about politics. As for Grade Pre-1, the degree of difficulty of the prompt type is similar to Grade 1. However, it provides some scaffoldings called *points* that may help examinees come up with their answers.

Levels for lower grades seem to vary in difficulty depending on to what extent the question is personal. The prompt in Grade 2 asks examinees’ opinions about a non-personal topic such as the food waste from restaurants and supermarkets. On the contrary, the prompt in Grade 3 asks about personal experience and can be answered without background knowledge.

**Setting of the Writing Tasks.** As for the settings of writing tasks, levels are constructed with instructional language, text length, and time allotment. Instructions are provided in English for the upper three levels while in Japanese for the lower two levels. As for the suggested text length, it varies from 25-35 words for Grade 3 to 200-240 words for Grade 1. Given that all grades are paper-and-pencil instead of computer-based writing, Eiken does not seem to differentiate levels based on test formats.

**Table 3**

Writing Tasks in Eiken (adapted from <https://www.eiken.or.jp/eiken/exam/>)

	Prompts and instructions provided in the writing test administered in January 2021	Suggested length	Instruction
1	<p><i>Topic:</i> “Are economic sanctions a useful foreign-policy tool?”</p> <p>Write an essay on the given TOPIC. Give THREE reasons to support your answer. Structure: introduction, main body, and conclusion</p>	200- 240 words	English
Pre-1	<p><i>Topic:</i> “Agree or disagree: Big companies have a positive effect on society”</p> <p>Write an essay on the given TOPIC. Use TWO of the POINTS below to support your answer. [Points: products, the economy, the environment, work-life balance] Structure: introduction, main body, and conclusion</p>	120- 150 words	English
2	<p><i>Topic:</i> “It is often said that restaurants and supermarkets should try to reduce the amount of food that they throw away. Do you agree with this opinion?”</p> <p>Write your opinions about the topic and provide two reasons to support your answer. <i>Points</i> are provided to help you with your writing. You may write from other perspectives. [Points: Cost, health and safety, and the environment] When you do not answer the <i>topic</i>, your writing may not receive a grade. Please read the topic carefully.</p>	80- 100 words	Japanese
Pre-2	<p><i>Question:</i> “Do you think it is a good idea for people to have a car?”</p> <p>You are asked about the <i>question</i> by your foreign friend. Write your opinions about the <i>question</i> and provide two reasons in English. When you do not answer the <i>question</i>, your writing may not receive a grade. Please read the topic carefully.</p>	50- 60 words	Japanese
3	<p><i>Question:</i> “Where do you like to go shopping?”</p> <p>You are asked about the <i>question</i> by your foreign friend. Write your opinions about the <i>question</i> and provide two reasons in English. When you do not answer the <i>question</i>, your writing may not receive a grade. Please read the topic carefully.</p>	25- 35 words	Japanese with Furigana (i.e., Japanese reading aid)



### **Scoring Validity**

Scoring validity is related to “all the aspects of the testing process that can impact on the reliability of test scores” (Shaw & Weir, 2007, p. 143). It can be established when scoring procedures such as the scoring guide accurately reflect the writing construct being measured (Weigle, 2002).

For the evaluation of writing performance in Eiken, a similar rubric consisting of four criteria is used at all five Grades: Grade 3

([https://www.eiken.or.jp/eiken/exam/2017scoring\\_3w\\_info.html](https://www.eiken.or.jp/eiken/exam/2017scoring_3w_info.html)), Grade Pre 2

([https://www.eiken.or.jp/eiken/exam/2017scoring\\_p2w\\_info.html](https://www.eiken.or.jp/eiken/exam/2017scoring_p2w_info.html)), and the other three upper

levels ([https://www.eiken.or.jp/eiken/exam/2016scoring\\_w\\_info.html](https://www.eiken.or.jp/eiken/exam/2016scoring_w_info.html)). The criteria include

*content*, *organization*, *vocabulary*, and *grammar*. *Content* pertains to clearness, persuasiveness, and concreteness of the content. *Organization* is judged with the effective use of expressions such as conjunctions. *Vocabulary* is evaluated with accuracy of spelling and meaning for the lower two levels (i.e., Grade 3 and Grade Pre-2) and the appropriate use of words and a variety of words for the upper three levels (i.e., Grade 2, Grade Pre-1, and Grade 1). *Grammar* is about grammatical accuracy and a variety of sentence patterns for all levels except for Grade 3, which is only evaluated with grammatical accuracy. Each criterion has zero to four points and thus the full score of a writing section is sixteen. Grade 1 has the same four criteria but each criterion has eight points so the full score is 32. The score is then calculated to a CES score.

The rubric for writing assessment used in Eiken is similar to the one used for an independent writing task in other standardized language tests. A rubric for an independent writing task in IELTS, for example, has criteria of *task response*, *coherence and cohesion*, *lexical resource*, and *grammatical range and accuracy*. The distinctive feature of Eiken is, however, its adoption of an analytic scoring approach. It allows raters to assign separate scores to each criterion in a writing rubric. The strength of this scoring approach is its detailed information about a test taker’s performance (Weigle, 2002). On the other hand, TOEFL and IELTS employ a holistic scoring approach in which raters provide only one score for overall writing (Crusan, 2013). Because of the cost- and time effectiveness, holistic scoring has been preferred for large-scale and high-stakes testing settings.

The reliability of a scoring approach seems to vary depending on writing task types. Ohta et al. (2018) found that a holistic approach yields higher reliability for integrated writing tasks

while earlier studies concluded that an holistic scoring approach was more reliable for independent writing tasks (Barkaoui, 2007; Schoonen, 2005). Since Eiken only has an independent writing task for all Grades, holistic scoring might be more appropriate to adopt. In terms of the scoring approach for the lower grades, however, analytic scoring would be more beneficial given that the lower grades are utilized as an achievement test at school. When the result is used for an educational purpose, detailed information about writing performance would help test takers develop their writing skills. Currently, STEP only reports a total score that encompasses all four criteria. To make the test more instructional, the separate scores in each dimension should be reported so that the examinees can use them as feedback.

### *Consequential Validity*

The validity of tests should also be considered from the consequential aspects including “evidence and rationales for evaluating the intended and unintended consequences of score interpretation and use in both the short- and long-term” (Messick, 1996, p. 251). Because the consequence of the test results is more significant for Grade Pre-1 and Grade 1, this section focuses on the validity of the test use for the top two grades.

According to CSE score, Grade 1 is equivalent to TOEFL 95-120, as well as IELTS Band 7.0-8.0. Grade Pre-1 corresponds to TOEFL 72-94 and IELTS 5.5-6.5. When writing sections in Eiken Grade 1 and Grade Pre-1 are compared to TOEFL and IELTS, which were specifically developed for university admissions, Eiken seems less demanding in the following three points. Firstly, Eiken has only one independent writing task (i.e., examinees write their response to a statement or a question) while the other two tests have two different tasks with two different rubrics. TOEFL, for example, has an integrated writing task in addition to an independent test. The integrated writing task requires test takers to read or listen to a certain passage and synthesize the information into their writing. This test is more demanding compared to the other in that it entails additional skills other than writing skills. Given that college students write a paper based on lectures and readings, an integrated writing task reflects real-world situations in an academic environment more than an independent writing task. Eiken’s writing section with only an independent writing task, thus, may not be able to infer test takers’ ability to write in an academic environment accurately.

Secondly, the suggested word length in the Eiken writing section is much shorter than in TOEFL and IELTS. Eiken requires 200-240 in Grade 1 and 120-150 words in Grade Pre-1 while TOEFL and IELTS suggests at least 300 words and 250 words respectively. Since TOEFL and IELTS have two sets of writing questions, examinees of the tests write more than 500 words. Although Eiken tries to measure examinees' writing skills to structure an essay with introduction, body, and conclusion, the short amount of writing would not likely be enough to achieve an accurate measurement.

Thirdly, all writing sections in Eiken are provided with hand-writing. This can also be a limitation to infer test takers' writing performance in academic settings. Given the growing exposure to technology, handwriting skills may thwart fair and valid assessment of writing (Barkaoui & Knouzi, 2018). If the results of higher levels such as Grades 1 and Grade Pre-1 are used as a language certificate for college entrance, it should also consider the implementation of computer-based testing. When it comes to the use of Eiken for university admissions, thus, the writing task format should be reexamined. It should take into consideration real-world needs that examinees would face in academic environments.

### **Conclusion**

Since its foundation in 1963, Eiken has endeavored to accommodate test structures to socio-temporal demands in the globalized era. It added two more grades between Grade 3 and Grade 2, and Grade 2 and Grade 1, and also introduced writing sections to lower grades. Thanks to the flexible changes in its format and scoring system, the test has now been used not only as an achievement test but also as a proficiency test for universities in Japan and overseas.

This paper reviewed the validity of writing sections across grades. Writing sections have a gradual increase in their linguistic demands and task settings. However, in terms of consequential validity, upper grades still need to be revised. To make the two grades more reliable and valid, the length of words, the task type, and the writing mode (handwriting vs. computer-based writing) should be reexamined.

## REFERENCES

- Bailey, K. M. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing*, 13(3), 257–279.  
<https://doi.org/10.1177%2F026553229601300303>
- Barakaoui, K., & Knouzi, I. (2018). The effects of writing mode and computer ability on L2 test-takers' essay characteristics and scores. *Assessing Writing*, 36, 19–31.  
<https://doi.org/10.1016/j.asw.2018.02.005>
- Brown, J. D., Davis, J.M., Takahashi, C., & Nakamura, K. (2010). Upper-level Eiken examinations: Linking, validating, and predicting TOEFL iBT scores at advanced proficiency Eiken levels. *Eiken Foundation of Japan*. Retrieved from  
<https://www.eiken.or.jp/eiken/group/result/pdf/eiken-toeflibt-report.pdf>
- Chapelle, C. A. (2012). Conceptions of validity. In G. Fulcher & F. Davidson (Eds.), *The Routledge handbook of language testing* (pp. 21–33). Routledge.
- Chujo, K., & Oghigian, K. (2009). How many words do you need to know to understand TOEIC, TOEFL & EIKEN? An examination of text coverage and high frequency vocabulary. *Journal of AsiaTEFL*, 6(2), 1–231.
- Crusan, D. (2013). Assessing Writing. In A. J. Kunnan (Ed.), *The companion to language assessment* (pp. 201–215). John Wiley & Sons, Inc.  
<https://doi.org/10.1002/9781118411360.wbcla067>
- Guapacha Chamorro, M. E. (2022). Cognitive validity evidence of computer- and paper-based writing tests and differences in the impact on EFL test-takers in classroom assessment. *Assessing Writing*, 51, 100594. <https://doi.org/10.1016/j.asw.2021.100594>
- Hamada, A. (2015). Linguistic variables determining the difficulty of Eiken reading passages. *JLTA Journal*, 18(0), 57–77. [https://doi.org/10.20622/jltajournal.18.0\\_57](https://doi.org/10.20622/jltajournal.18.0_57)
- Japan National Tourism Organization. (2022). The examination guidelines for the national translators (lit. Zenkoku Tsuyaku Annaishi Shiken).  
[https://www.jnto.go.jp/jpn/projects/visitor\\_support/interpreter\\_guide\\_exams/operation\\_2022.pdf](https://www.jnto.go.jp/jpn/projects/visitor_support/interpreter_guide_exams/operation_2022.pdf)
- Kachru, B. B. (1985). Standards, codification and sociolinguistic realism: The English language in the outer circle. In R. Quirk & H. G. Widdowson (Eds.), *English in the world:*

- Teaching and learning the language and literatures* (pp. 11–30). Cambridge University Press.
- Kane, M. (2010). Validity and fairness. *Language Testing*, 27(2), 177–182.  
<https://doi.org/10.1177/0265532209349467>
- Kobayashi, H., & Rinnert, C. (2002). High school student perceptions of first language literacy instruction: Implications for second language writing. *Journal of Second Language Writing*, 11(2), 91–116. [https://doi.org/10.1016/S1060-3743\(02\)00067-X](https://doi.org/10.1016/S1060-3743(02)00067-X)
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13(3), 241–256. <https://doi.org/10.1177/026553229601300302>
- Miura, T., & Beglar, D. (2002). The Eiken vocabulary section: An analysis and recommendations for change. *JALT Journal*, 24(2). <https://doi.org/10.37546/JALTJJ24.2-1>
- Ohta, R., Plakans, L. M., & Gebril, A. (2018). Integrated writing scores based on holistic and multi-trait scales: A generalizability analysis. *Assessing Writing*, 38, 21–36.  
<https://doi.org/10.1016/j.asw.2018.08.001>
- Piggin, G. (2011). An evaluative commentary of the grade 1 EIKEN test. *Language Testing in Asia*, 1(4), 144–167. <https://doi.org/10.1186/2229-0443-1-4-144>
- Plumb, C., & Watanabe, D. (2016). A critique of the Grade 2 Eiken test reading section: Analysis and suggestions. *Shiken*, 20(1), 12–17.
- Shaw, S. D., & Weir, C. J. (2007). *Examining writing: Research and practice in assessing second language writing*. Cambridge University Press.
- Schoonen, R. (2005). Generalizability of writing scores: An application of structural equation modeling. *Language Testing*, 22(1), 1–30. <https://doi.org/10.1191/0265532205lt295oa>
- Society for Testing English Proficiency. (n.d.-a). Test structures of each grade.  
<https://www.eiken.or.jp/eiken/exam/>
- Society for Testing English Proficiency. (n.d.-b). Writing assessment policy  
<https://www.eiken.or.jp/eiken/exam/eiken-scoringquality.html>
- Society for Testing English Proficiency. (n.d.-c). Eiken criteria  
<https://www.eiken.or.jp/eiken/exam/criteria/>

Weigle, S. C. (2002). *Assessing writing*. Cambridge University Press.