

A critique of the Grade 2 EIKEN test reading section: Analysis and suggestions

Christopher Plumb¹ and Daita Watanabe²

cplumb@adnoc.ae, watanabe-d@shitennoji.ed.jp

1. Abu Dhabi National Oil Company Technical Institute

2. Shitennoji Junior/Senior High School

Abstract

The increased use of the EIKEN tests from a nationally used junior and high school proficiency test within Japan to a broadly used academic entrance test both in Japan and internationally means that the test requires more research and documentation than is currently available. By utilizing the EIKEN Grade 2 Reading test as an example, this paper argues that there are limitations in the test's construction, validity and documentation. The paper also briefly argues that the test suffers partially from a "washback" effect, which is related to both the test's construction and its use as an overall entrance test with only a broadly defined construct validity. The paper concludes that much more internal documentation from the test's producers is required as well as independent research from outside individuals and institutions to verify to tests overall utility.

実用英語技能検定(英検)が日本国内の中高の英語能力テストから、日本そして国際的な入学試験として使用されることは、現在以上にテストについて研究と検証を必要とすることを意味する。例として英検2級のリーディングを用いることで、この論文はテスト作成、妥当性そして検証に制限があることを論じている。また簡単に、英検が washback 効果(テストが指導に与える影響)から部分的に悪い影響を受けていることを論じている。その washback 効果はテスト作成、そして大きく定義されるテスト作成の妥当性のみ備える入学試験としての活用に関連している。この論文は英検テストの有益性を明らかにするために、第三者による個人そして外部組織による研究と同時にテスト作成者による内部検証が必要だと結論付けている。

Keywords: EIKEN, reading assessment, validity, washback effect, high-stakes testing

The Eigo Kentei or EIKEN test is Japan's most popular and widely administered test of English language proficiency (Eiken, 2016a). Supported directly by the Japanese government, it was created by the Society for Testing English Proficiency, now known as the Eiken Foundation of Japan (Eiken), in the early 1960's and has been used by all Japanese prefectures in the public education system as a benchmark. Recently some colleges and universities have used the test for adults as a placement standard at the international level (Eiken, 2016a). It is designed to be used as an alternative to the internationally popular TOEIC and TOEFL tests, but initially designed specifically for Japanese junior and high school students (Tamura, 2006). According to Eiken, the test is widely available in Japan, is offered at less than half the cost of comparable tests, and features secure administration and score reporting (Eiken, 2016b; Tamura, 2006). The test has also begun to be more readily accepted outside of Japan, as demonstrated in the case in Australia where it was initially only accepted in Queensland but later nationally (Muroko, 2014). Due to the recent expansion of the test beyond Japan's borders, Eiken has attempted to produce more English language documentation about the test's design, administration and methodology; however, little has been published so far.

The EIKEN test system has seven levels (although these are confusingly reported as five levels with two sublevels); this review will focus on the reading section of the "Grade 2" test developed for the 2015(2) year (three versions are released each year). The test is divided into four main components: listening, reading, writing, and speaking. Grammar is often regarded as a fifth component and integrated into the other sections (Nakanishi, Hayashi, Kobayashi, & Sakuma, 2010; Tamura, 2006). The Japanese Ministry of Education, Culture, Sports, Technology (MEXT) uses the Grade 2 test as a benchmark for English proficiency at the high school level upon graduation (MEXT, 2002). As using the Grade 2 test in this manner is one of the most widely used applications of the EIKEN test system, the reading component will

be critiqued from within this context. In addition, one of us is a Japanese high school teacher who has considerable experience with the EIKEN Grade 2 test. In particular, this paper will examine the reading section's format, the validity and washback effects of the test and the documentation of the Grade 2 test and EIKEN testing system in general.

Reading Section Format and Tasks

The reading test, a sample of which is available online (Eiken, 2015), is divided into four sections consisting of different reading tasks, most often requiring the test taker to engage in "cloze" exercises. Section one is comprised of 20 cloze-based short answer questions; there is a choice between four multiple choice answers, all of which belong to the same grammar constructs (adverbs for example). Section two consists of five questions where students must choose the correct word order from a spread of five words. Section three has two sub-sections, where students must select the proper answer to complete a cloze exercise within paragraphs. In one sub-section there are four choices for each question, each belonging to the same grammar construct. In the other, there are four choices, some of which belong to different constructs. Section four differs from the previous sections; it assesses comprehension of short written passages and asks contextual questions. Similar to the third section, it is divided into sub-sections. The first analyzes an e-mail message and asks three multiple-choice questions, while the remaining two contain a short article and four or five multiple-choice questions.

Aspects of the test format and tasks present themselves as targets for criticism. Though the section is designed to assess reading as a whole, almost two-thirds of the test concentrate on grammar-based questions, and do not focus on understanding meaning or what could be defined as overall comprehension. Research has shown that reading comprehension is a difficult concept to assess (Koda, 2004); however, some aspects of the test could be focused more towards assessing overall meaning by using different question types. In addition, if the test is used as a benchmark for Japanese high school students, then some of the questions could be perceived as being out of a high-school context. For instance, questions 16 and 17 relate to situations long time employees at companies would encounter (Eiken, 2015, p. 3). One could argue that question 17 is simply testing the phrasal expression "bottom line" rather than a context, but surely this can be constructed in a manner more relevant to a "real world" situation a high school student might encounter. Another case in point is the sample e-mail. The email discusses the parameters and requirements for a security company to move its office to a new location. This sample email represents an irrelevant situation for the average high-school student and should have been developed through the use of different subject matter (Eiken, 2015, p. 7). Though Eiken insists that the tests must be relevant to the test takers at each level, this is not always the case, or at the minimum is difficult to measure (Eiken, 2016b). As little has been documented about the exact nature of the design and format of the test, it is difficult to theorize about the test's design and how its questions are developed and verified.

Validity and Documentation

According to Cumming (2012), a second language examination with a high validity rating necessitates that the outcome of the test depend solely on its construct for assessment and not other points. Therefore, a test's validity requires a necessary framing of the way it conceptualizes and addresses what is language competence. This makes it difficult to discuss the validity of the EIKEN Grade 2 test, because the test has not been adequately documented. Additionally, Piggin (2011) notes that because Eiken does not adequately demonstrate what it defines as language ability and states that the test is to be used as a "broad spectrum of language ability", its construct validity is questionable. While Eiken has begun the process, it has not completed collecting the data for the construct in order to undertake a comprehensive and cohesive validity study (Eiken, 2016b). In addition, few outside researchers have studied the validity or reliability of the EIKEN test. One of the few to do so studied the EIKEN 1 test and noted that outside

research into the test has been limited and concludes that much more is required (Piggin, 2011). Sarich (2012) noted the same and went further, stating that because in practice the EIKEN test is often utilized for many different uses to measure proficiency (both inside and outside Japan), the test is not always being used for the purposes it has been designed for; as documentation related to the tests design is incomplete, this makes use of the test for different tasks more problematic. What outside research that has been done often focuses on comparisons between EIKEN Grade 1 test scores and TOEIC, TOEFL, and CEFR benchmarks (Dunlea & Matsudaira, 2009; Eiken, 2016b). Little research has been done on the other four levels, or specifically on their subsections such as reading. Additionally, and perhaps most concerning is the documentation listed by Eiken itself, which includes many references, but very few after 2008 aside from those that compare the test to the aforementioned benchmarks (Eiken, 2016b).

As part of constructing the EIKEN test's validity, Eiken published the EIKEN "Can-do List" revealing what each grade level should be able to perform, on the presumption that they are engaged in, "English in real-life situations" (Eiken, 2008, p. 4). Beginning in 2003, Eiken spent three years consulting with over 20,000 test takers to develop the Can-do list (Eiken, 2016c). In terms of reading at the Grade 2 level, the list states that students who pass the test should be able to "understand lengthy expository texts and find necessary information in texts of a practical nature" (Eiken, 2008, p. 10). Therefore, a passing score on the Grade 2 test implies the ability to read guidebooks for travelers, follow practical traffic directions, understand newspapers with Japanese explanations in footnotes, recognize the central argument of texts, comprehend sales pamphlets and easily distinguish between the topic and support sentences when reading paragraphs (Eiken, 2008). The Can-do lists represent the backbone of the EIKEN tests' benchmarking, yet very little independent research outside of the Eiken organization has been attempted (Dunlea, 2010; Nakanishi, et al., 2010). Additionally, some of the tasks referenced in the can-do lists are vague, such as stating a student proficient at the grade 2 level could "understand expository texts written for a general audience". The tasks are provided without any rationale as to what linguistic features they require or what band descriptors they may match with related standards, such as the CEFR.

In one such inquiry with students and their English instructors, Nakanishi, et al. (2010) investigated the validity of the Can-do list. They found contradictions between the list and the students' actual performance; students often could not replicate in practice the Can-do list's requirements. For instance, many learners could not accurately understand train schedules. It was theorized that this was because learners in Japan are not allocated much time for instruction focusing on authentic communicative tasks. Similarly, in anecdotal evidence provided by Piggin (2011), students commented that the test was too structured on precise details such as vocabulary, or that they could pass the test by studying test-preparation books, but the test itself was not an accurate reflection of their real abilities. Though in some cases English instructors could connect criteria taught in the classroom with the criteria of the Can-do list, the list does not discuss grammar with any criteria regarding syntax accuracy, in spite of the fact that the test, especially the reading section, contains significant sections devoted to grammar (Nakanishi, et al., 2010). Nakanishi, et al. (2010) argued the list should be adjusted to a model that considered the tasks and activities which commonly form the real curricula in Japanese classrooms. Indeed, since MEXT uses the Grade 2 EIKEN examination as a benchmark for proficiency, research should be done not only from within Eiken itself but also through various academic institutions (MEXT, 2002).

Washback

Though defined multiple ways, washback as a concept has been concisely defined as, "the impact of a test on teaching" (Wall & Alderson, 1993, p. 41). Research suggests that washback can have positive and/or negative effects on classroom instruction. For example, in Japan, teachers and students tend to focus their attention on the demands of a test, especially if that test is considered to be a high-stakes test (Buck, 1988, cited in Bailey, 1999; Piggin, 2011). This means a focus is often placed on test-taking skills, reducing

instruction on more communicative tasks. Test taking skills may include distinguishing between multiple-choice distractors or scanning for contextual clues. While it is common knowledge that Japanese English teachers often “teach to the test” with respect to the high-stakes testing - they teach specific skills that may be required for a specific test - little attention has been paid to the effects of washback concerning the EIKEN tests (Sarich, 2012).

Despite the limited research on washback effects related to EIKEN tests, there is some evidence that negative effects in the classroom have occurred as a result of its format. The reading section in particular can be singled out for such criticism. Tamura (2006) noted there was a “strategic approach to the EIKEN grammar” and that teaching to the test should be considered a reasonable strategy to pass the EIKEN test. With more than half of the reading section of the Grade 2 test being comprised of grammar-based questions, the approach seems practical. However, the effect on overall English study in the classroom could be problematic as resources and time are shifted to studying grammar instead of other areas that might improve students’ overall English ability. It should be noted, however, that any effects of washback directly from EIKEN tests could be considered as much a problem of overworked instructors trying to assist their students with their future without the benefit of understanding the nature of how the tests are produced and what they are designed for. If this anomaly is a product of instruction related to taking the test, rather than simply a product of the test design itself, and is surely in need of much further research to determine its effect on both the test results and overall instruction in classrooms in Japan.

On 2 March 2009, it was reported that Ikubunkan Middle and High School in Tokyo had been illegally coaching students to pass the EIKEN tests after seeing the questions in advance. The teacher who opened the sealed envelopes in advance held special classes before the test to teach the students exactly how to pass (RakutenBlog, 2009). This represents a rare and drastic example of washback. Nonetheless, it demonstrates the lengths teachers are willing to go to prepare their students for the EIKEN tests. The high-stakes nature of the test as well as the way it is administered can be assumed as being partially responsible for the result. Eiken acknowledges that the EIKEN test may contribute to washback in the classroom, but has stated that proving the existence and effects of washback are difficult to ascertain with certainty (Eiken, 2016b). Finally, though Eiken states the test should test “English in real-life situations” (Eiken, 2008, p. 3), Piggen notes in her EIKEN Grade 1 study that the test is not reflective of real-world language and tends to focus more on the “professional world of work” (2011, p. 152) meaning that instruction to pass the test may focus in this area. This observation regarding the Grade 1 test can be applied to some of the material in the Grade 2 reading test (Eiken, 2015). This focus on professional English may have something to do with the test sometimes being used in the Japanese private sector as a means to assess English language ability for possible career advancement (Sarich, 2012). If so, it would mean that the test is not fully designed for its initial stakeholders (high school students for example) and that may mean that preparation for the test is done specifically to pass the test’s components, rather than to increase overall English ability.

Limitations and Further Research

The most striking limitation of the EIKEN tests is the lack of documentation with respect to the test’s validity construct. Eiken has acknowledged this limitation and has started to develop studies and commission research to document the test correctly (Eiken, 2016b). A serious concern that must be taken into consideration is that the EIKEN test has seven levels (5 main with 2 sub-levels in grades 1 and 2), that are divided into four sections. Research must be initiated to look at all aspects of each of the grade levels and each of their sections. The Can-do List has provided a reasonable start to establish the validity and reliability of the test, but as noted in the Nakanishi, et al. (2010) study it requires significant revision. It also requires research to determine what skills and linguistic requirements tasks listed in the Can-do statements are required to achieve success. Relating these to international standards such as the CEFR could

possibly make the test appear much more transparent. Additionally, especially as the EIKEN tests are being utilized more and more abroad, though Eiken has taken the initiative to begin to produce more research and documentation, due to its large scale use and often high-stakes nature, research from the institution that develops the test is insufficient. Independent research from other institutions, test developing bodies and academic specialists in the field needs to be done to corroborate findings Eiken have already disseminated and intend to in the future. Both Piggin (2011) and Sarich (2012) noted this several times in their studies. Finally, research into what possible effects washback may have in the classroom should be undertaken and above all, shared with instructors preparing students for the tests.

Conclusion

The EIKEN test has grown from a low-stakes national test to assess English proficiency into a high-stakes test, supported by the government in the public education system and used by international bodies for entrance into colleges and universities. This massive growth has not been in conjunction with research developed on the test. The developers of the test are aware of this fact and have started to implement some change and research. Once research has been completed, it can be reviewed both internally and by external experts to determine if either revision or more research is needed. As gathering validity evidence is a time consuming and ongoing process, producing new, widespread research on such a heavily utilized test is of high importance. In addition, research on other aspects of the test, such as its administration and the effects of washback on classroom instruction should be addressed as well. At present, other high-stakes tests which the EIKEN compares itself to have had significantly more research and development put into them, especially through the work of outside, independent researchers. More importantly, what limited research that has been conducted on the EIKEN system itself has been placed on evaluating the Grade 1 tests that compare similarly to the TOEFL test. The time has come for the same detailed efforts to be put into the Eigo Kentei tests at all levels.

References

- Bailey, K., M. (1999). *Washback in language testing*. Princeton, NJ: Educational Testing Service.
- Cumming, A. (2012). Validation of language assessments. In C. Chapelle (Ed.), *The encyclopedia of applied linguistics*. Malden, MA: Blackwell Publishing Ltd.
- Dunlea, J. (2010). The EIKEN Can-do List: Improving feedback for an English proficiency test in Japan. In L. Taylor & C. J. Weir (Eds.), *Language testing matters: Investigating the wider social and educational impact of assessment - Proceedings of the ALTE Cambridge Conference April 2008* (pp. page numbers needed). Cambridge, UK: Cambridge University Press.
- Dunlea, J., & Matsudaira, T. (2009). Investigating the relationship between the EIKEN tests and the CEFR. In N. Figueras & J. Noijons (Eds.), *Linking to the CEFR levels: Research perspectives*. Arnhem, NL: Cito, Council of Europe, and EALTA.
- Eiken. (2008). The EIKEN can-do list: English translation Retrieved from http://www.eiken.or.jp/eiken/exam/cando/pdf/Eiken_CandoList_translation.pdf
- Eiken. (2015). EIKEN test: Grade 2. Retrieved from https://www.eiken.or.jp/eiken/exam/grade_2/pdf/201502/2015-2-1ji-2kyu.pdf
- Eiken. (2016a). History and purpose Retrieved 20 February, 2016, from <http://stepeiken.org/history-and-purpose>
- Eiken. (2016b). Research Retrieved 20 February, 2016, from <http://stepeiken.org/research>

- Eiken. (2016c). 英検 Can-do リスト Retrieved 20 February, 2016, from <http://www.eiken.or.jp/eiken/exam/cando/>
- Koda, K. (2004). *Insights into second language reading: A cross-linguistic approach*. Cambridge, UK: Cambridge University Press.
- MEXT. (2002). Japanese government policies in education, culture, sports, science and technology 2002 Retrieved 20 February, 2016, from http://www.mext.go.jp/b_menu/hakusho/html/hpac200201/hpac200201_2_015.html
- Muroko, M. (2014, 23 October). Australia accepts Eiken certificate as proof of English ability, *The Japan Times*. Retrieved from <http://www.japantimes.co.jp/news/2014/10/23/national/australia-accepts-eiken-certificate-as-proof-of-english-ability/#.VsmjYpx97mF>
- Nakanishi, C., Hayashi, C., Kobayashi, W., & Sakuma, S. (2010). Development of can-do statements for college students : A preliminary study. *大学英語教育学会(JACET)関東甲越地区研究年報 (JACET Kanto Region Annual Report)*, 6, 14-23. Retrieved from http://ci.nii.ac.jp/els/110008664789.pdf?id=ART0009744109&type=pdf&lang=en&host=cinii&order_no=&ppv_type=0&lang_sw=&no=1456055406&cp=
- Piggin, G. (2011). An Evaluative Commentary of the Grade 1 EIKEN Test. *Language Testing in Asia*, 1(4), 1-24. doi: 10.1186/2229-0443-1-4-144
- RakutenBlog. (2009). School admits leaking Eiken info since 1995 Retrieved 20 February, 2016, from <http://plaza.rakuten.co.jp/odrix/diary/200903020000/>
- Sarich, E. (2012). Accountability and external testing agencies. *Language Testing in Asia*, 2(1), 26-44. doi: 10.1186/2229-0443-2-1-26
- Tamura, A. (2006). The significance of the obtainment of the certificate in the English proficiency test (Eiken), and a proposal of an effective and efficient learning and teaching method to successfully prepare for passing: A strategy for internalizing basic Eiken grammar. *釧路工業高等専門学校紀要 (Kushiro Technical College Research Reports)*, 40, 45-48. Retrieved from <http://www.kushiro-ct.ac.jp/library/kiyo/kiyo40/tamura40.pdf>
- Wall, D., & Alderson, J. C. (1993). Examining washback: the Sri Lankan Impact Study. *Language Testing*, 10(1), 41-69. doi: 10.1177/026553229301000103