

# Introducing Kyoto Appeal: Issues in and Implications of Using Four-Skills Proficiency Tests as Entrance Examinations in Japan

Atsushi MIZUMOTO  
*Kansai University*

## Abstract

This work introduces the Kyoto Appeal, a statement jointly issued by three major Japanese academic associations in the field of applied linguistics and English teaching in September 2013. The Kyoto Appeal addresses concerns and suggestions related to the use of four-skills proficiency tests for university entrance examinations in Japan. The present paper discusses issues and implications related to the implementation of Kyoto Appeal focusing on three perspectives: (a) washback, (b) construct representation, and (c) practicality. Based on theoretically grounded and expert views, thorough consideration must be exercised in using four-skills proficiency tests for university entrance examinations.

## 1. Background

Four-skills English proficiency tests have received ever-increasing attention in Japan in recent years. In April 2013, the Japan Association of Corporate Executives (*Keizai Doyukai*) released the proposal ‘Achieving University Entrance Exams that Test Practical English Abilities’, in which they called for the use of the Test of English as a Foreign Language Internet-based Test (TOEFL iBT) as the university English language entrance examination test. In July 2014, the Headquarters for Revitalization of Education, an advisory body to the Prime Minister, submitted a reform proposal for university entrance examinations, suggesting the use of four-skills proficiency tests, such as TOEFL and International English Language Testing System (IELTS). Along with these proposals, the Japanese education ministry intends to overhaul the university entrance examination administered by the National Center to make the English test more communicative and measure all four skills.

These movements have led academic organisations to formulate their professional views on the issue of using four-skills proficiency tests in entrance examinations in Japan. This collection is known as the Kyoto Appeal.

## 2. Kyoto Appeal

The Kyoto Appeal (<http://www.jasele.jp/2013/09/17/kyoto-appeal-2013/>) was jointly announced by three major Japanese academic associations in the field of applied linguistics and English teaching, namely Language Education and Technology (LET), the Japan Society of English Language Education (JASELE), and the Japan Association of College English Teachers (JACET), in September 2013. The official title of the document is ‘On a reform proposal for university entrance examinations submitted by the Headquarters for Revitalization of Education’. As its official title suggests, the Kyoto Appeal was produced as a reaction to the suggestion from the governmental panel regarding entrance examinations; as the first document of its type in this field in Japan, the Kyoto Appeal reflects the professional views of the three abovementioned academic associations. As each association has approximately 2,000 to 3,000 members, the release of this document was considered appealing and influential. Below is an independent (unofficial) translation.

---

### **Kyoto Appeal**

‘On a reform proposal for university entrance examinations submitted by the Headquarters for Revitalization of Education’

#### **A) Premises**

We recognise the importance of reform in the current English entrance examination system. However, the reform should consider the English as a Foreign Language (EFL) context in Japan, and discussions should be based on accumulated scientific data.

#### **B) Ideal characteristics of alternative English entrance examinations**

- (1) The test should assess the achievement of high school English language education based on the National Curriculum Guidelines. In addition, test scores should be used for the purpose of selection.
- (2) The test must be one that measures four skills, including speaking and writing, following the National Curriculum Guidelines.
- (3) The test must have a positive washback effect on improving teaching practice and forming learning abilities in high school.
- (4) The test should demonstrate satisfactory reliability, validity, practicality, and fairness in the assessment of English language abilities of the target test takers (high school students).
- (5) The test should, if possible, provide multiple test-taking opportunities.
- (6) The same test system should be adopted for subjects other than English.

### **C) On using commercial tests such as TOEFL**

To meet the criteria suggested above, the admission office of universities should select a type of selection procedure suitable (e.g., the so-called Admission Office entrance examination) and unsuitable (e.g., the general entrance examination) for using commercial tests. Further, we take a flexible view on the use of a commercial test as a certification examination. Nonetheless, the test should go through validation studies to confirm that it assesses the English language abilities of Japanese high school students reliably and validly.

### **D) Proposals**

- (1) As the first phase, the Ministry of Education, Culture, Sports, Science and Technology (MEXT) should revise the National Center Exam so that it can test four skills.
- (2) As the second phase, MEXT should play a leading role in creating a new test that reflects the National Curriculum Guidelines (as in the Korean National English Ability Test or NEAT).
- (3) In the process of developing a new test, the research findings of related fields, such as second language acquisition, applied linguistics, e-learning (ICT), and language testing, should be used as references and utilised fully. Additionally, to evaluate the reliability, validity, practicality, and fairness of the test (see Section B-4 of this document), MEXT should employ professional knowledge of educational measurement and assessment.
- (4) The Language Education and Technology (LET), Japan Society of English Language Education (JASELE), and Japan Association of College English Teachers (JACET) shall actively participate in the development of a new test.

### **Postscripts**

- (1) A commercial test such as TOEFL may be considered to be required at graduation from university, depending on the needs and educational goals of faculties at each university.
  - (2) It is strategically important for Japan to promote foreign language education not limited to English language education and implement reforms related to learning other foreign languages in high school.
  - (3) The pros and cons of post-hoc disclosure of all test items, as currently conducted, should be reconsidered, especially in view of creating an item bank (i.e., accumulation of items) and reusing tests.
  - (4) We shall likewise give due consideration to students with disabilities and take appropriate measures to ensure that their right to take the test is protected.
-

### 3. Issues and implications

The Kyoto Appeal has suggested that using four-skills proficiency tests for entrance examinations in Japan has its merits. It also advocates the development of a new test aligned with the National Curriculum Guidelines rather than using commercial tests, which may have potential consequences. This section presents a discussion of the issues and implications of using four-skills proficiency tests for Japanese entrance examinations in general, and implementing the Kyoto Appeal in particular. The inclusion of four-skills tests in Japanese entrance examinations brings a number of possible benefits, challenges and consequences; the focus here will be on three perspectives: (a) washback, (b) construct representation and (c) practicality.

#### 3.1 Washback

Washback, or the influence of language testing on teaching and learning (Cheng, Watanabe and Curtis, 2004), has been researched in the field of language testing since the 1990s. Studies on washback show that it is a complex process. Under the premise that introducing a four-skills English test to the entrance examination system in Japan will cause a positive washback, then a theoretical update of English language teaching and testing in Japan can be expected to ensue. That is, an increasing number of stakeholders would appreciate the current view on “the nature of language ability and language use” (Bachman, 2000) reflected in the test items, as McNamara (2013) suggested:

*The domain of language tests, and the resulting test construct, reflect theories of language in linguistics, of language use in sociolinguistics and linguistics pragmatics, of language processing in psycholinguistics, and of language learning in the field of language acquisition research. Developments in each of these fields have had successive impacts on the design and validation of language tests. (p. 341)*

In second or foreign language teaching, the communicative approach has been dominant for decades. Approaches such as task-based language teaching (TBLT), content and language integrated learning (CLIL), and project-based learning (PBL) have come to the forefront and increasingly gained attention among practitioners. Likewise, communicative language testing, in which integrative performance on a given task, most likely composed of four skills, has been the most accepted approach.

Meanwhile, only receptive skills, namely, reading and listening (often reading only), are tested using multiple-choice items in the current English entrance examination in Japan. Writing is rarely included, and speaking, never. This scenario is attributed to a practicality problem (see Section 3.3). Four skills and their related subcomponents can be further divided into ‘knowledge’ and ‘performance’, and thus, multiple-choice test items for receptive skills would mostly tap into linguistic knowledge. By including a four-skills proficiency test or creating a new four-skills test,

as the Kyoto Appeal suggests, Japan can finally catch up with the up-to-date theoretical underpinnings of language assessment and then sufficiently measure the integrative, productive skills of test takers (i.e., writing and speaking).

Reconceptualisation of English teaching and learning as a result of introducing a four-skills test might be possible because teachers and learners can infer the current view on the nature of language ability and language use by looking at the test items. As reported in Watanabe (1996), which investigated the effect of the university entrance examination on teaching, teachers' beliefs lead to the prevalent and persistent use of the grammar-translation teaching method in Japan. In the 20 years since Watanabe's study, the situation has not changed much (Takagi, 2010). In turn, learners' beliefs on English language learning are formed in such a context. Moreover, textbooks and entrance examinations have been aligned with the grammar-based syllabus. As a result, teachers teach and learners learn English 'as a school subject', heavily dependent on the teaching, learning, and testing of linguistic knowledge. Performance aspects are rarely given attention. A top-down introduction of four-skills tests in Japanese university English language entrance examinations has the potential to make stakeholders reconceptualise the *de facto* standards of English teaching and learning.

Another positive consequence would be improvement on the test items in English entrance examinations. English entrance examinations in Japan have long been criticised for problems in terms of their "purposes, effects, roles, responsibilities, and perspectives" (see Brown, 2002). After a series of criticisms, Brown (2002) candidly argued, "The future of Japanese young people is far too important to be left in the hands of the hundreds of teams of amateur test developers in the many universities all over Japan" (p. 102). If the current English entrance examination system can be changed as the Kyoto Appeal suggests, then those test items that can cause negative washback on teachers and learners could be eliminated totally or improved to an extent.

The above are the positive consequences of introducing a four-skills proficiency test or creating a new four-skills test. However, these are lay people's views, to be blunt. As studies on washback generally show, the reality is not as simple. Washback studies on the introduction of new tests in education systems (e.g., Wall and Alderson, 1993) mostly suggest that "things do not work as intended; there is no simple relationship between the use of a test and its effects" (Fulcher, 2010, p. 278).

Over the years, washback research has amassed evidence that, even after the introduction of new tests, teachers often stick to their way of teaching. Watanabe (2004) summarised the situation as follows: "The examination does not influence teachers' fundamental beliefs and attitudes about teaching and learning, the role of teachers and students, and how teaching and learning should be carried out" (p. 130). Therefore, to produce positive washback from innovation in testing, much work is needed for the large-scale retraining of, and support for, teachers.

In addition, communication among stakeholders is necessary to engender positive changes on test use. Although language testing experts have gathered crucial evidence on test use through

decades of washback studies, they have not sufficiently communicated their findings with potential audiences (O’Sullivan, 2015). Further, the low literacy in language assessment among stakeholders is another concern for the introduction of new tests. The tester is thus responsible for reaching potential audiences and addressing issues in the use and interpretation of test scores.

In introducing a four-skills test to the entrance examination system in Japan, the intention is to bring about positive and much needed changes in English language education in the country. Nevertheless, it is simply wrong to expect all positive changes to happen by introducing a new test; negative consequences of the current English language education in Japan (e.g., low scores in proficiency tests, students being barely able to speak or write English, or low motivation) cannot be explained by only tests. Issues such as a ridiculously large class size for language teaching and learning and short contact hours to acquire a language have to be taken into consideration as well. These critical issues cannot be replaced by the introduction of a more desirable test (i.e., a four-skills test); such an idea is based on an optimistic, unrealistic, and often naïve view that tests can change everything.

### **3.2 Construct representation**

Introducing a four-skills proficiency test or a new four-skills test based on the National Curriculum Guidelines to entrance examinations needs the careful examination of constructs measured with those tests. As the interpretation and use of test scores (i.e., the validity of the scores) differ considerably depending on a purpose, scores on a test can be “highly valid for one purpose and much less valid for another” (ETS, 2015). Thus, test constructs must match the test purpose.

In examining the issues of using four-skills tests in English entrance examinations within the framework of “Assessment Use Argument” (Bachman and Palmer, 2010), Imao (2013) pointed out that “clearly defined construct[s]” are missing in the discussions on using four-skills tests. In other words, it is not clear what should be realistically measured for university applicants in Japan. In the case of TOEFL iBT, it measures listening, reading, speaking and writing skills necessary in an academic environment when students study abroad. A number of TOEFL iBT validation studies have been conducted, and the investigation of its factor structure indicates that it measures the four skills as intended (Sawaki, Stricker and Oranje, 2009; Gu, 2014). However, the construct “representativeness” comes under serious question when TOEFL iBT is used for the selection of students for university admission in Japan. Similarly, developing a new four-skills test based on the National Curriculum Guidelines, as proposed in the Kyoto Appeal, is a challenging task because constructs represented in the National Curriculum Guidelines are too broad and vague. Thus, constructs need to be defined based on the target language use of Japanese high school graduates when they enter university (Imao, 2013).

Indeed, the defined constructs come only from the items in a test; therefore, no test is perfect for all purposes. In a review of studies reporting factor structures of proficiency tests, Harsch (2014) argued as follows:

*One has to concede that such studies look into language proficiency via the “window” of test scores derived from specific test instruments. The scores, at least in part, are thus influenced by the construct that is reflected in the test structure...At least in part, the research findings seem to be influenced by the instruments we use to look at language as well as by the methods of analysis. (p. 154)*

This statement attests to the fact no matter the reliability and validity of a test, a careful review of constructs is necessary to investigate the match between the test constructs and purpose.

Defining constructs in the paradigm of communicative language testing grows increasingly challenging because communicative language testing is more concerned with tasks and test content. Thus, as Fulcher (2010) noted, “The evaluation of performance in a communicative language test [is] done purely on the basis of behavioural outcomes, defined as the degree to which the test taker achieve[s] the intended communicative effect” (p. 113). The Common European Framework of Reference (CEFR) (Council of Europe, 2001) was built upon this type of ‘performance models’ in stark contrast to ‘construct models’, upon which the field of language testing has long depended (e.g., Bachman, 1990). Although CEFR and its can-do statements, which are anchored on performance models, have theoretical problems in terms of reference to language competence (see Fulcher, 2010), the concept has made its way through educational contexts worldwide, including Japan.

With CEFR gaining momentum in many educational settings, proficiency test developers have sought to map test scores or grades according to CEFR standards. One recent example of such an attempt in Japan is the Common Scale for English produced by EIKEN (<https://www.eiken.or.jp/cse/>). Although such a move is commendable as it provides a practical guideline for test users to compare test scores, comparisons of constructs represented by each test are completely left out. A careful examination of the constructs by a panel of testing experts is, therefore, necessary if four-skills proficiency tests are used for university entrance purposes.

### **3.3 Practicality**

Bachman and Palmer (2010) defined practicality as “the difference between the resources that will be required in the development and use of an assessment and the resources that will be available for these activities” (p. 262). In terms of test practicality, introducing a four-skills proficiency test or developing a new four-skills test faces few concerns and challenges. If four-skills proficiency tests are used for entrance examinations, the cost of developing, administering, scoring and giving results to test-takers is not an issue because these tests are commercially available to anyone. However, test fees tend to be high for students, especially when



tests are used for university entrance admission, given that such tests are only for English among other subjects in entrance examinations.

Meanwhile, developing a new four-skills test aligned with the National Curriculum Guidelines, as the Kyoto Appeal suggests, involves all the above practicality concerns and challenges. Additionally, maintaining a newly developed testing system will entail considerable costs. The Kyoto Appeal suggests following the footsteps of Korea in administering, in which four skills are tested. However, the Ministry of Education decided to abolish the plan to develop and use NEAT as a college entrance English language test. The decision to scrap the plan is mostly attributed to the large budget for developing and maintaining NEAT. As Green (2014) stated, “an assessment that lacks resources will not be sustainable. If there is not enough time, equipment, money or expertise to operate an assessment, it may never be used and certainly will not survive in use for very long” (p. 59). From NEAT’s case, the lesson is that test practicality should be carefully considered to put the Kyoto Appeal into action.

Nonetheless, practicality should not be overly prioritised as to compromise validity. Weir (2005) warned as follows: “We should not consider method before trait. Practicality considerations are often allowed to intrude at too early a stage and validity is often threatened rather than enhanced as a consequence” (p. 49). Thus, a balance needs to be struck between practicality and validity in four-skills test development.

#### **4. Final comments**

This paper introduced the Kyoto Appeal along with the issues and implications related to its implementation. The Kyoto Appeal was publicised when public attention mounted toward using four-skills proficiency tests for entrance examinations in Japan. Although the inclusion of four-skills tests in university entrance selection has certain positive effects, as the Kyoto Appeal advocates, the issues and implications addressed in this work are rarely discussed. All stakeholders, educators or not, have a say in English tests (mostly based on personal and subjective views), and often, tests are used as a tool at various levels of politics (Shohamy, 2001). Thus, greater professional participation is needed in public concerns and discussions of language test use. The Kyoto Appeal is significant in that it is an expression of the shared determination of three major Japanese academic associations to help discuss the use of four-skills proficiency tests in entrance examinations, drawn from professional expertise. As such, I would like to emphasise the importance of supporting the Kyoto Appeal and keeping raising professional voices from language testing experts.

Only time will tell if the Kyoto Appeal is nothing more than words on paper or a page-turning moment in the history of English language assessment in Japan (e.g., Sasaki, 2008).



## Acknowledgments

This work is based on the author's talk at the Second British Council Symposium on Supporting Reform in Education: New Directions English – Role of English Assessment in Internationalisation held at Meiji Kinenkan on 30 September 2014. I would like to thank the steering committee members of JASELE for giving me an opportunity to present the Kyoto Appeal at the conference and in this work on behalf of JASELE. The ideas and opinions expressed in this paper, except the Kyoto Appeal itself, are the author's and do not reflect those of JASELE or other organizations.

## References

- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford University Press.
- Bachman, L. F. (2000). Modern language testing at the turn of the century: Assuring that what we count counts. *Language Testing*, 17, 1–42. doi:10.1177/026553220001700101
- Bachman, L. F. and Palmer, A. S. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford University Press.
- Brown, J. D. (2002). English language entrance examinations: A progress report. *Curriculum Innovation, Testing and Evaluation: Proceedings of the 1st Annual JALT Pan-SIG Conference*, 95–105. Retrieved from <https://jalt.org/pansig/2002/HTML/Brown.htm>
- Cheng, L., Watanabe, Y. and Curtis, A. (Eds.). (2004). *Washback in language testing: Research contexts and methods*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge University Press.
- ETS. (2015). Glossary of standardised testing terms. Retrieved from [https://www.ets.org/understanding\\_testing/glossary/](https://www.ets.org/understanding_testing/glossary/)
- Fulcher, G. (2010). *Practical language testing*. London, UK: Hodder Education.
- Green, A. (2014). *Exploring language assessment and testing: Language in action*. New York, NY: Routledge.
- Gu, L. (2014). At the interface between language testing and second language acquisition: Language ability and context of learning. *Language Testing*, 31, 111–133. doi:10.1177/0265532212469177
- Harsch, C. (2014). General language proficiency revisited: Current and future issues. *Language Assessment Quarterly*, 11, 152–169. doi:10.1080/15434303.2014.902059
- Imao, Y. (2013, September). Examining the issues in the English portion of university entrance exams in Japan using the assessment use argument (AUA) framework. *Building an argument*

- for language assessment use in Japan*. Symposium conducted at the 17th Annual Conference of JLTA (Japan Language Testing Association), Waseda University, Tokyo.
- McNamara, T. (2013). Language testing: History, validity, policy. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, J.-I. C. Hansen, N. R. Kuncel, S. P. Reise and M. C. Rodriguez (Eds.), *APA Handbook of Testing and Assessment in Psychology, Vol. 1: Test theory and testing and assessment in industrial and organizational psychology* (pp. 341–352). Washington: American Psychological Association. doi:10.1037/14047-021
- O’Sullivan, B. (2015). Validity, validation and development: Building and operationalising a comprehensive model. *JLTA (Japan Language Testing Association) Journal*, 18, 25–33.
- Sasaki, M. (2008). The 150-year history of English language assessment in Japanese education. *Language Testing*, 25, 63–83. doi:10.1177/0265532207083745
- Sawaki, Y., Stricker, L. J. and Oranje, A. H. (2009). Factor structure of the TOEFL Internet-based test. *Language Testing*, 26, 5–30. doi:10.1177/0265532208097335
- Shohamy, E. (2001). *The power of tests: A critical view of the uses of language tests*. Harlow, UK: Pearson Longman.
- Takagi, A. (2010). *A critical analysis of English language entrance examinations at Japanese universities* (Doctoral dissertation). Retrieved from <https://ore.exeter.ac.uk/repository/bitstream/handle/10036/117893/TakagiA.pdf>
- Wall, D. and Alderson, J. C. (1993). Examining washback: The Sri Lankan impact study. *Language Testing*, 10, 41–69. doi:10.1177/026553229301000103
- Watanabe, Y. (1996). Does grammar translation come from the entrance examination? Preliminary findings from classroom-based research. *Language Testing*, 13, 318–333. doi:10.1177/026553229601300306
- Watanabe, Y. (2004). Teacher factors mediating washback. In L. Cheng, Y. Watanabe and A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 129–146). Mahwah, NJ: Lawrence Erlbaum Associates.
- Weir, C. J. (2005). *Language testing and validation: An evidence-based approach*. Basingstoke, UK: Palgrave Macmillan.