

Proposing change in university entrance examinations: A tale of two metaphors

David Allen

allen.david@ocha.ac.jp

Ochanomizu University

Abstract

This article describes a recent education reform initiative concerning English education in Japan, specifically the proposed introduction of four-skills tests as part of the university entrance admissions process. The first aim is to summarize, in English, some of the key issues and events concerning the reform. To this end, background information and a timeline of key events since 2016 is provided. The second aim is to contrast proposals made by two academic organizations, the Japan Language Testing Association (JLTA) and the Science Council of Japan: Language and Literature Committee (SCJ). It is shown that, while agreeing on a number of specific issues related to the reform, these two organizations take starkly different positions in terms of their general orientation, which, it is argued, reflects the background of the organizational members and their views on foreign language education in Japan. These contrasting positions are discussed with reference to the metaphor, to throw the baby out with the bathwater. Finally, it is argued that a number of criticisms levelled at the proposed use of private four-skills tests illustrate a reluctance to engage with issues related to the currently used university entrance exams; in other words, these criticisms are made while ignoring the elephant in the room.

Keywords: university entrance exams, four-skills tests, MEXT, communicative language education reform

Educational systems are complex. They involve, at one level of abstraction, a *curriculum* (what is to be taught and learned), the *delivery* of the curriculum (teaching, materials, learning environments), and *assessment* of learning (tests, evaluation methods and materials; see Bunch, 2012; O’Sullivan, 2020). These three elements (curriculum, delivery, and assessment) must work together for a system to function optimally. When they do not, there is a problem.

In Japan, a key issue concerns the assessment in the English education system. The assessment comes at the end of the national course of study (NCS), which for many students functions as a stepping stone to further education¹. The problem is that the content of the curriculum and that of the assessment do not align. The Ministry of Education, Culture, Sports, Science, and Technology (hereafter, MEXT) curricula through elementary, junior high and senior high school are designed to help students develop knowledge of the English language and the ability to *both* comprehend and produce language in *both* spoken and written modes (MEXT, 2017a, 2017b, 2018a). The final assessment, which functions as an achievement test, is the Common Examination for University Admission (*daigaku nyūgaku kyōtsū tesuto*; hereafter, the Common Test). The Common Test was introduced in 2020 with the first English test section to be administered in January 2021. Prior to the Common Test, English was assessed in the National Center Test (NCT, *sentā shiken*; see Watanabe, 2013). The Common Test has a reading and a listening section, and is essentially a continuation of the NCT. It is in multiple-choice format and the reading and listening texts include both written and spoken discourse genres. While many, including myself, see the Common Test as a significant improvement on the NCT², the fact remains that neither test directly assesses productive abilities in English, as they do not require the candidate to write or speak. In other words, there is a fracture in the system created by the difference in curriculum and assessment.

An additional issue is that of the university entrance exams (UEEs)³. These UEEs include the second-stage exams administered by national and public universities (i.e., *nijishiken*), which applicants must take in addition to the Common Test (or prior to 2020, the NCT). The UEEs also include the in-house developed entrance exams of private universities, which are typically taken by applicants in addition to, or in place of, the Common Test/NCT, though some universities require only the Common Test/NCT scores (see Kuramoto & Koizumi, 2018). Content analyses of UEEs have shown that they typically test reading ability, with a focus on grammatical and lexical knowledge, and regular use of translation tasks, especially in national/public university exams (Brown & Yamashita, 1995a, 1995b; Kikuchi, 2006; Watanabe, 1997). Given that these tests are typically perceived as high-stakes tests that are taken by candidates immediately after completion of the high school curriculum, they can impact what is taught and learned in schools (e.g., see Green, 2014) and test preparation contexts, such as cram schools (e.g., Allen, 2016b). Consequently, not only is the assessment of productive skills lacking in the Common Test, it is also lacking, either completely or in terms of balance, in the UEEs. This situation creates a conflict for high school teachers and learners: Should they follow MEXT's balanced curriculum, or should they instead focus on the knowledge and abilities needed to navigate the assessments?

In August 2016, MEXT outlined a reform in which the English section of the Common Test administered by MEXT is to be abolished in 2024, at which point universities should adopt one of a number of four-skill tests administered by private companies to assess the English ability of candidates (MEXT, 2016). During the transitional period of 2020 through 2023, universities have three options: They can utilize the common test only, a four-skills test only, or a combination of tests. This upheaval of the articulation process between pre-tertiary and tertiary education has created considerable anxiety and confusion among stakeholders. Commentators have taken to mainstream and social media, generating a storm of criticism and opinion. Consequently, in late 2019, MEXT postponed the initiative (MEXT, 2019) and convened a new committee to find a solution. This committee has been in session throughout 2020 (MEXT, 2020c) and MEXT has been soliciting opinions from a wide range of stakeholders (MEXT, 2020b).

Below is a non-exhaustive annotated timeline of some key events beginning in 2016 that are particularly relevant to this reform initiative. More extensive background and historical information on the various reforms and proposals concerning English language education in Japan can be found elsewhere (e.g., Butler & Iino, 2005; Kuramoto & Koizumi, 2018; Otsu et al., 2013; Sasaki, 2008)⁴. Importantly, the majority of the events in the timeline below were held, and books and articles published, in Japanese. Therefore, for non-Japanese speakers, it is difficult to keep up-to-date and informed on the issue. Hence, one aim of this paper is to fill this gap by presenting the views of academics in Japan who are experts in language assessment or a related field (e.g., linguistics, literature, education, educational measurement)⁵. By doing so, English-language readers can hopefully better understand not only the problems and their proposed solutions, but also the apparently conflicting views on the direction of language education in Japan.

Timeline of events related to the reform

- Aug 2016: MEXT outlines proposal to introduce four-skills private tests (MEXT, 2016)
- Sep 2016: Japan Language Testing Association (JLTA) Annual Conference, themed ‘Between Validity and Practicality of University Entrance Exams Based on Communication Skills: The Possibility of Reform’ http://jlta2016.sakura.ne.jp/?page_id=18
- Jan 2017: JLTA (2017) proposal submitted to MEXT
- Dec 2017: Abe (2017) ‘*English education in chaos...*’ is published
- Mar 2018: The University of Tokyo, Center for Research and Development on Transition from Secondary to Higher Education (*Tōkyōdaigaku kōdai setsuzoku kenkyū kaihatsu sentā*) Symposium <https://www.ct.u-tokyo.ac.jp/news/20180210-symposium/>
- Apr 2018: Koizumi (2018) ‘*How to choose and use four-skills tests ...*’ is published
- Jun 2018: Haebara (2018) edited collection is published
- Aug 2018: The Japan Association for Language Learning and Technology (LET) Symposium, themed ‘Reconsidering four-skill assessment of foreign language proficiency’ http://www.j-let.org/let2018/page_20180222024053
Symposium at the Japan Association of College English Teachers (JACET) international convention, themed ‘Current and future assessment of four skills in an entrance examination’ <http://www.jacet.org/convention/2018-2/>
- Mar 2019: The University of Tokyo Center for Research and Development on Transition from Secondary to Higher Education (*Tōkyōdaigaku kōdai setsuzoku kenkyū kaihatsu sentā*) Symposium (2) <https://www.ct.u-tokyo.ac.jp/news/20190110-symposium2019/>
- Aug 2019: Symposium sponsored by the National English Education Society (JASELE, *Zenkoku eigo kyōiku gakkai shusai no shinpojiumu*), themed ‘Symposium on the significance and issues of the English four-skills assessment in the university entrance examination’ <http://www.jasele.jp/symposium2019/>
- Sep 2019: Symposium ‘Evaluating fairness and justice of university entrance English examinations in Japan’ at JLTA Annual Conference http://jlta2016.sakura.ne.jp/?page_id=606
- Oct 2019: Education Minister Koichi Hagiuda creates uproar with his gaffe about university applicants competing ‘in accordance with their standing’ (The Mainichi, 2019)
- Nov 2019: Official proposal of using four-skill tests for university entrance purposes from 2020 postponed (MEXT, 2019)
- Dec 2019: Committee formed to discuss future direction (MEXT, 2020c)
British Council New Directions Conference, themed ‘Realising Potential: Policy, Engagement and Impact’ <https://www.britishcouncil.jp/en/new-directions/about/theme>
- Aug 2020: Science Council of Japan (2019) proposal is published
MEXT collects public opinions (669 submissions received in one-month period) (MEXT, 2020b)
- ...
- Apr 2024: Planned implementation of the new system

Introducing two proposals

MEXT's proposal to use four-skills English tests as part of general route university admissions has met with resistance from many stakeholders; even those in favor have typically held some reservations about its implementation.⁶ What is perhaps most interesting, however, is the emergence of a clear difference in position that can be seen by analyzing two formal proposals made by academic organizations in Japan in response to the reform. These two proposals are the primary focus of this paper.

The first is by the Japan Language Testing Association (2017, henceforth, JLTA), released in Japanese and English, and entitled *Proposal for handling English testing within the 'Prospective university entrance scholastic abilities evaluation test [provisional name]'*. The contents were submitted to MEXT as a position statement in January, 2017. The nine authors include Yoshinori Watanabe (president of JLTA), Rie Koizumi, and other well-known language assessment experts based in Japan. The position of the JLTA committee is made clear from the outset: 'While endorsing their [MEXT's] general orientation, the JLTA herewith voices our opinions on the range of issues regarding the policy based on our expertise.' (p. 1). In other words, the JLTA committee took the position that MEXT's aim of improving language education in Japan through improved alignment of the curriculum, teaching, and assessment, is a commendable one. Nevertheless, given the number of concerns raised by the JLTA committee, there are several serious problems concerning its conception and implementation. Therefore, the JLTA committee proposed a series of strategies and requirements for arriving at evidence-based solutions to the issues of test quality and implementation feasibility. Specifically, the committee recommended mobilizing language testing researchers to assist in these evaluations, the results of which should be made transparent and in accordance with state-of-the-art research and practice in the field. They also called on the test agencies to ensure certain requirements are met and evidence for meeting them is to be provided. Moreover, they suggested that individual universities conduct their own systematic studies to determine which, if any, test is appropriate for their purposes. In sum, this position agrees with the initiative to move towards more comprehensive assessment of language skills, both receptive and productive, in line with the NCS, while it also details serious issues with the innovation. In other words, *let's keep the baby, but throw out the bathwater*.

The second proposal was published by the Science Council of Japan: Language and Literature Committee (2020, henceforth, SCJ), released in Japanese and entitled '*Recommendations for English exams of university entrance exams*.' This committee consists of thirteen academics in Japan, including Takane Ito (chair) and Yoshifumi Saito (vice chair), both at the University of Tokyo. The committee is comprised of many regular commentators on the issue, including Kumiko Torikai and Yukio Otsu. In addition, Masahiko Abe and Tomokazu Haebara, both regular commentators, were present in an advisory capacity. The position taken by the SCJ committee is clearly more critical not only of the details of the specific proposal but also the general move towards assessment of four skills. Firstly, the proposal states that given the limited input available in the EFL context of Japan, learners must acquire explicit (not implicit) knowledge of the English vocabulary and grammar system, and that it is crucial to use the first language (Japanese) and the written English language in doing so. This 'basic knowledge' (*kiso chishiki*), which underlies all language use, according to the SCJ, cannot be developed through activities focused on the four skills. Secondly, it argues that given that language learning proceeds from comprehension (i.e., via receptive skills) to production, aiming for a balance of four skills is inappropriate. Therefore, learners

must acquire receptive knowledge sufficiently before developing productive skills, which entails that the curriculum, and assessment, should focus primarily on the former. They argue that attempting to focus on productive skills to achieve a balanced proficiency will, in fact, impede learners' overall growth in the future. All in all, these statements reveal that the SCJ's position runs contrary to the four-skills approach adopted in the NCS, and in fact, represents what could be termed a more traditional view of language education. In addition, the SCJ committee lists numerous problems with the proposed innovation, which make it impossible in practice to assess four skills at the national level as part of the Common Test. They suggest instead that ultimately the decision, that is, whether to use a four-skills test for entrance purposes or not, should be up to each university. In sum, the SCJ committee's proposal outlines reservations concerning both the communicative approach to curriculum and assessment, suggesting that the aim of communicative reform is unnecessary, and indeed, unwanted. In other words, *throw out the baby along with the bathwater*.

Specific issues raised in the proposals

The main problems identified in MEXT's innovation are described below. Most have been taken up in some detail in books, symposia, and committee meetings noted in the above timeline, and so are only discussed briefly here. It should also be noted that many of the problems were identified relatively early and test agencies were requested to submit information that was made available online (4skills.jp), though this website is no longer in operation.

Both the JLTA and the SCJ proposals point out that MEXT's recommendation to utilize a wide range of tests (i.e., Cambridge English Assessment, EIKEN, GTEC, IELTS, TEAP, and TOEFL; a total of 23 tests when all levels are included) creates serious problems for test users in terms of determining which test is appropriate. This is because each test differs in its intended purpose, its intended target test-taker and the target proficiency. With the aim of facilitating appropriate selection and use of tests, the JLTA recommended disseminating information on these facts to ensure all test users can select and use tests appropriately.

Both proposals note the limitations of the Common European Framework of Reference (CEFR; Council of Europe, 2001) conversion table (MEXT, 2018b) for comparing scores of tests that have different purposes, content and score ranges. The SCJ also registers the concern that the CEFR levels are determined by the test companies themselves, not by third-parties; moreover, some tests have changed CEFR benchmarks, revealing instability that would lead to confusion (p. 6). The SCJ also note the can-do descriptors are not detailed criteria that can be used for assessing candidate ability in language assessments (see Weir, 2005, for early criticisms on this point) and they were not designed for the purpose MEXT is intending (i.e., to select students based on score comparisons across tests). They also criticize MEXT's claim that the CEFR is an 'international indicator' (p. 7).

Both proposals discuss the alignment (or lack thereof) of the tests with the NCS. The JLTA request that evidence be provided demonstrating the extent of alignment between each test and the NCS. Moreover, both proposals note the suitability of the four-skills tests for individual universities' needs. The JLTA proposal states that 'it is desirable that each university engages in systematic and specific studies, and selects and uses tests found to be appropriate.' It also refers to 'needs analysis in language education and

test creation.’ (p. 3), which is noteworthy because the needs analysis is a cornerstone of communicative language teaching syllabus design (e.g., Munby, 1978). In contrast, the SCJ proposal suggests that university exams should measure candidate abilities with respect to the curriculum in the university, its admission policy and its educational philosophy; and therefore, each university should determine whether or not to utilize a specific, or indeed any, four-skills test.

Both proposals stress the likelihood of unequal opportunities to take private four-skills tests due to economic and regional disparities in terms of location of test centers and the frequency of testing. That is, not only are the tests themselves expensive, but also candidates who live in rural areas may need to pay for travel and accommodation, further increasing their financial burden. In addition, students with disabilities may be disadvantaged, especially when it comes to speaking tests. Although MEXT has collected information on test agencies’ abilities to address these concerns, they remain perhaps the most controversial aspect of the innovation.

Both proposals registered concern about the information that is made publicly available about each test, though the wording differs notably: The JLTA proposal reads, ‘Testing agencies must publish the kind of detailed information that is essential for selecting tests... test’s purpose and targeted proficiencies, scoring criteria and methods, as well as how the test was drafted and implemented and methods for its appropriate use.’ (p. 2). The SCJ proposal states that ‘the quality of questions and other issues regarding fairness are left to test companies and the actual situation is unknown,’ and that ‘grading criteria for speaking and writing are sometimes vague and unspecified’ (p. 6).

Only the JLTA specifically refers to washback (i.e., the effect of a test on teaching and learning; see pp. 3-4). It notes ‘the introduction of comprehensive four-skills testing *by itself* does not *necessarily* lead to improvements in English education at the high-school level’ (italics added). It also recognizes that while appropriate test use can generate a positive effect, inappropriate use can generate a negative effect, such as narrowing of the curriculum. It notes the importance of ‘engineering a positive effect of motivating test-takers to work towards improvement in four skills.’ Furthermore, it recommends that to support positive washback from a test, test agencies must provide a score report and feedback that is instructive for guidance and learning. In contrast, the SCJ document does not once mention washback (*hakyū kōka*). It does state, however, that because four-skills tests differ in purpose from the NCS, they will end up replacing the current curriculum (p. 5), which appears to be an expectation of negative washback. Moreover, only the JLTA proposal stresses the importance of teacher training, which must be strengthened in terms of programs and content, ‘so it may fully reflect the results of language testing research’ (p. 2). Essentially, this emphasizes the importance of developing teachers’ assessment literacy. The SCJ proposal does not mention teacher training directly, though states the need to improve school education and contents, and to strengthen English education at universities.

Only the SCJ registers concern that maintaining confidentiality and responding to unforeseen circumstances are overseen solely by the test agencies. Moreover, only the SCJ argues that language cannot simply be divided into four skills. Further, because neither the new companion volume to the CEFR (Council of Europe, 2018) nor the NCS does this in practice, it is not necessary to assess these skills individually, and doing so would be inconsistent with the NCS.

A general observation can be made at this point: In line with the general orientations of the two proposals, while the JLTA's concerns are accompanied by concrete suggestions for mitigating them, the SCJ's concerns are made with the aim of supporting rejection of the proposal. This is perhaps partly due to the time difference: The JLTA made its proposal within six months of MEXT's announcement, while the SCJ proposal came four years after it. Moreover, the SCJ proposal was released when MEXT had already formally postponed the plan and was holding a committee to determine the future course of action. Nevertheless, the difference in approach is also indicative of underlying theoretical positions on the nature of language learning in the Japanese context, which dictate the general orientation towards, or away from, four-skills assessment. This issue is discussed in the following section.

Discussion

Here, my intention is to point out what I see as a fundamental difference in orientation of the two proposals, and by extension the different perspectives of academics in Japan who have commented on the entrance exam issue. To this end, I refer more broadly to the literature but try to stay focused on views of authors and researchers who are connected to the proposals, primarily by their affiliation, but also, by extension, their academic background (i.e., language assessment researchers and applied linguists, or those in related disciplines, such as linguistics, education, or literature). Firstly, I overview relevant research in language assessment, then I examine the commentary of the SCJ authors. Finally, I take up the second metaphor in this article's title, the elephant in the room.

As noted above, the JLTA proposal adopts a solution-oriented, research-informed approach to the problem. A similar approach is evident in a number of recent publications by those involved in language assessment. Perhaps most notably is Koizumi (2018), which is a 263-page book entitled '*How to choose and use four-skills tests: From the viewpoint of validity.*' The aim of the book is clear from the title; however, Koizumi's aim is also to develop the reader's assessment literacy by overviewing test validity frameworks and validation studies. As noted above, this is in line with the principles of the JLTA, and other language testing organizations in Japan, such as the JALT TEVAL SIG. In another study, Baba (2019) sought to find solutions to the problem of how to generate positive washback in the high school context through the application of theory in educational psychology. He notes the lack of a solution-oriented approach in the wider discussion: 'With the exception of very few cases (e.g., Koizumi, 2018), the academic discussion has completely ignored the issues of how to reduce negative washback and whether, and how, positive washback can be generated' (p. 45). It is pertinent to note that Baba's observation is also true of the SCJ proposal, which does not once refer to washback. Most recently, Allen (2020) argued that research into the use of four-skills tests in Japan is necessary to determine, on the basis of evidence, whether specific uses of tests in specific contexts can be deemed appropriate. Concrete suggestions for future studies were provided, such as domain and needs analyses at universities, investigations into curriculum alignment, and impact studies. In sum, these studies all share a common aim, which is to resolve the issues at hand while maintaining a focus on improving the educational system through learning, teaching and assessment of all four skills. Again, they seek to *keep the baby, throw out the bathwater*.

In addition to the above studies are those that directly investigate test washback. Although some early washback studies were conducted in Japan, such as Watanabe (1996, 1997), over the last few years, there has been a marked increase in the number of studies that have sought to understand the impact that

introducing four-skills tests can have in the Japanese educational context (e.g., Allen, 2016a, 2017; Allen & Nagatomo, 2019; Green, 2014; Sato, 2018, 2019; Nagatomo & Allen, 2019). One aim of these studies has been to see how a specific test influences learning and/or teaching in a specific context. Another aim is to understand the washback mechanism at a general level so that it can be better predicted in specific educational contexts. From an assessment perspective, washback and impact studies can be considered paramount to predicting and understanding the effect of tests in educational contexts.

Now, we turn to the work of a number of authors present in the SCJ committee who continue to be vocal critics of MEXT's various communication-oriented reforms. A good example of this can be found in Otsu, Erikawa, Saito and Torikai's (2013) book entitled *English education in danger*. All of the authors argue from different perspectives for the importance of the 'basic abilities' (*kiso-ryoku*) of grammar, vocabulary, and reading. For instance, Otsu (2013) makes his view clear: 'The problem with emphasizing "communication" in English education policy is the deemphasizing of grammar' (p. 65). These authors view MEXT's various reforms as threats to English education in Japan and view themselves as its protectors. These threats include the use of four-skills tests for entrance purposes, conducting English classes in English, and introducing English into primary schools. To hammer home the message, this is depicted visually on the front cover of the 2013 text as four characters (the authors) attempt to deflect these incoming attacks.

In other work, Torikai (2018, and see Fujiwara, 2018, who summarized panelists' arguments at a regional JACET symposium) asserts that reading ability is the foundation of the four skills and should be the focus of teaching in schools. Consistent with this belief, she is also highly critical of the proposal to assess four skills. Her stance is that, 'in the university entrance examination, test-takers' basic English ability should be assessed through their reading skill' (Fujiwara, 2018, p. 2) and that, 'each university can develop comprehensive English ability after admission' (Torikai, 2018, p. 142). In other words, pre-tertiary education should focus on developing 'basic English ability' (*eigo no kiso-ryoku*), which is characterized as the ability to read and knowledge of sentence structure, grammar and vocabulary, while university education should develop students' productive abilities.

A related but somewhat more aggressive argument is that of Abe (2017). He singles out a number of proponents of four-skills tests (the so-called 'Neo-four-skills group', which includes two education ministers, a cram school teacher and author, and an academic advisor for the MEXT curriculum) and argues that they are not trying to reform the curriculum and assessment system for the benefit of the students, but for financial gain. That is, they are promoting an ideology of an oral approach to language education, which includes, for instance, the recommendation to teach English in English, and by privatizing the exam system, money will inevitably flow to test companies, materials providers and cram schools. Abe's own views on language teaching are distinctly vague and impressionistic yet converge with the idea that there must be a focus on the basics: Reading and listening should be primary; writing and speaking should be left until later. Consistent with this view, he sees no need to drastically reform the exam system.

The claims of the SCJ members above are broadly in line with the key points emphasized in the SCJ position statement: Learners must acquire explicit (not implicit) knowledge of the English vocabulary and grammar system (the basics); it is crucial to use Japanese in class (not English) and to focus on the written

(not oral) language; learners must acquire receptive knowledge sufficiently before developing productive skills; and focusing on productive skills will impede learners' future language development. In sum, there is a belief in the centrality of reading (i.e., the written mode, receptive ability) in pre-tertiary education, which contradicts the balanced communicative approach of the curriculum.⁷ The committee thus argues against the use of balanced four-skills tests; in other words, *throw out the baby and the bathwater*.

The elephant in the room

In this section, we turn to the second metaphor in the title. When combing through the criticisms levelled at private tests and test agencies, the discerning reader will note that many are often equally applicable, if not more so, to the entrance exams that are currently in use. This inconvenient truth, or the elephant in the room, is often omitted entirely from the discussion. In other words, critiquing private four-skill tests without applying the same critique to the current exams is conveniently, but inappropriately, ignoring the major problem (the elephant!). Some of these criticisms are discussed below.

The private four-skills tests may not match the NCS (noted in both proposals). Although it is clear that some tests (e.g., TOEFL) are less aligned with the NCS guidelines than others (e.g., EIKEN), this must be demonstrated through research as recommended by the JLTA. But what about the current university entrance exams? The NCT and Common Test are based directly on the NCS and therefore should align well. However, these tests do not test productive knowledge and ability, which is, in fact, the primary reason for reform. In other words, these tests do not align with the curriculum because they only assess a subset of it. Considering the UEEs, they also suffer from the same problem of imbalance in the skills assessed. Moreover, early content analyses of these exams have shown that they diverged considerably from the NCS in terms of text complexity and vocabulary range (e.g., Brown & Yamashita, 1995a, 1995b; Hasegawa, Chujo, & Nishigaki, 2006; Kikuchi, 2006). Their reliance on word re-ordering and translation tasks likely also renders them ineffective at identifying candidates with writing abilities specified in the NCS (Moore, 2015). In sum, even Torikai has noted, 'there is also a view that individual entrance examinations at some universities are already outside the curriculum guidelines' (Fujiwara, 2018, p. 2).

Assessing four skills individually is inconsistent with the NCS, which aims to nurture integrated language abilities (SCJ, p. 4). This claim insists that the assessment of four skills individually is theoretically problematic and is indicative of a general belief that MEXT's reform proposal is fundamentally flawed and should be rejected. However, putting aside the numerous possible counter-arguments to this claim, the same question must be asked of the current exams: How theoretically consistent are these tests with the NCS? According to the preceding discussion, it is unlikely that a strong argument could be made in their support.

It is impossible to measure productive skills ... because it is difficult to guarantee fairness (SCJ, p. 5); *the quality of questions and other issues regarding fairness are left to test companies and the actual situation is unknown* (SCJ, p. 6); *grading criteria for speaking and writing assessments are sometimes vague and unspecified* (SCJ, p. 6). These criticisms concern the reliability of subjective grading and the transparency of test procedures, and again, they are likely to apply to some tests/test agencies more than others. Considering the case of the Cambridge Assessment English and IELTS, for instance, numerous empirical research papers are available online that have examined the content and procedures of these tests (see

www.cambridgeassessment.org.uk and www.ielts.org), and public versions of grading descriptors are similarly available. Moreover, a glance through the accompanying reports for a relatively new test, the TEAP, reveals state-of-the-art test development procedures (see www.eiken.or.jp). In contrast, it is well-known that very little information is publicly available concerning the design and evaluation of the UEEs⁸. Many questions have been asked by researchers over 25 years ago (Brown & Yamashita, 1995c; O'Sullivan, 1995), though responses have not been forthcoming. For example, what is the intended purpose of the exam? What are the specific abilities targeted by items? What is the reliability? What grading criteria are used for scoring written responses (in English or Japanese)? How are raters trained? From a 'measurement' perspective (e.g., Brown, 1996; Kuramoto & Koizumi, 2018), these questions are legitimate and fundamental, certainly not taboo, and they should be addressed by test developers in a transparent fashion. It is true that Japan's so-called 'test culture' (Yoshida, 1996) and the 'principle of education' (Kuramoto & Koizumi, 2018) may explain some of the idiosyncrasies in approaches to testing in Japan. However, when tests are being proposed for the same test use situation, that is, for general admissions at universities in Japan, it is unacceptable to expect certain standards from one kind of test (i.e., four-skills tests) but not another (i.e., UEEs).

Introducing four-skills tests will turn secondary schools into test preparation schools (SCJ, p. 5). This criticism concerns the predicted washback of four-skills tests on school education. Implicitly, however, it assumes that the current exam system has no, or at least less, negative impact. Although determining what is negative impact is, to some extent, a matter of perspective, it is clear that the current entrance exams do have an impact on teaching and learning in schools both within the mainstream and shadow education sectors (e.g., Allen, 2016b; Green, 2014; Watanabe, 1996, 1997). Despite the fact that washback is undoubtedly a complex phenomenon, and the way teachers teach and the way students learn will never be completely dictated by a test, the content of current exams is perceived by teachers and learners to influence the content of their English teaching and learning. For example, see the self-report data of 190 university students in Allen (2017) and of 3,766 high school students and 390 teachers in Green (2014). The impact of UEEs is not a myth, or an unfounded belief, as claimed by some (e.g., Torikai, 2013, p. 97). Therefore, when critiquing the potential impact of newly introduced tests, academics must also be cognizant of the impact of those in current use.

Conclusions

There exists a problem with the Japanese educational system in terms of the English curriculum, its delivery and assessment. The proposal to introduce four-skills tests was intended to address the problem by making all aspects of the system fall into line. However, this proposal has been postponed and is now under review.

By comparing two proposals, this study has shown that while the JLTA and SCJ overlap in many of the specific criticisms of MEXT's proposal, they diverge in terms of their position on the general aim and hence their approach to the problem. The JLTA was shown to be supportive of the general aim of assessing four skills, presenting numerous concrete suggestions, including the deployment of language assessment researchers and professionals. In contrast, the SCJ was shown to be critical not only of the aim to assess four skills but also of the aim to teach four skills in a communicative, balanced approach adopted in the NCS.

The SCJ proposal appears to distance itself from the concept of test washback, perhaps because the impact of the current exams is more in line with their vision of what language study should involve (i.e., reinforcing a focus on primarily the written mode and receptive skills). Meanwhile, perhaps they are concerned about the potential impact of four-skills assessment because it is contrary to this vision. Conversely, the JLTA proposal indicates that engineering positive washback and mitigating negative washback is a core concern, in line with current approaches to language assessment, in which the consequences of using an assessment are (or should be) a test user's first concern (e.g., Koizumi, 2018).

It is these points of conflict which highlight the starkly different viewpoints of the two proposals. If academics have fundamentally different views on the curriculum, naturally they will have fundamentally different views on the assessment. In other words, if the SCJ position is taken that a balanced four-skills curriculum is undesirable, it is impossible to argue for the use of balanced four-skills tests. The alternative position, which appears to be adopted by the JLTA, is that it is not essential to change the curriculum. However, if there is to be a change in the assessment so that it better aligns with the curriculum, there needs to be even greater strengthening of the delivery, particularly regarding teacher training and teaching materials.

Unfortunately, the stark difference in opinion here is deep-rooted and reflects what Henrichsen (1989) observed: "The debate between those who advocate the 'practical' objectives of oral English study and those who defend the traditional 'cultural' purpose of language study, which is more closely allied with the nature of the examinations, is a long standing one and has not yet been resolved" (p. 178; cited in Watanabe, 2004, p. 132). This debate continues to affect both discussions of exams and the national curriculum, as demonstrated here in the two proposals.

At the time of writing, the future of the proposed innovation is still undetermined. It is unknown whether the baby will indeed be thrown out with the bathwater and whether the elephant will continue to go unacknowledged. For the sake of the millions of stakeholders in the Japanese English education system, especially the learners and the teachers, let us hope that those making the decisions will make fair and just ones, and that the English education system in Japan is remedied so that it functions as intended.

Acknowledgements

I am indebted to the numerous expert readers who have reviewed and commented upon earlier drafts of this article. Any errors or omissions remain entirely my own.

Notes

¹ Over recent decades there has been a rise in the number of admissions to university via alternative routes, such as recommendation-based, Admissions Office (AO) and special examinations; this is particularly the case for private universities, where around half of incoming students enter via these 'alternative' admissions (MEXT, 2020a). Consequently, not all students in Japan are required to take the final assessment in order to enter university and thus the discussion presented in this paper is most relevant to applicants who sit English examinations as part of their university entrance admission procedure.

² One example of a much-needed and now implemented revision is the removal of items that required test-takers to select the correct stressed syllable in a word (see Buck, 1989). Another is the inclusion of

listening texts that are heard only once, rather than twice, and the inclusion of more varieties of English (see Yanagawa, 2007).

³ In this paper, the Common Test and NCT are not included within the definition of UEEs, though these tests are indeed used for university admissions. The reason for the distinction is that the Common Test/NCT is developed centrally by MEXT as an achievement test and as an entrance test for any/all universities, while the UEEs are developed by individual institutions for their specific admissions purposes.

⁴ For brevity the time line begins in 2016; however, the background to the initiative stretches back many decades and includes countless relevant events that cannot be documented here. Reviewers, however, have noted that 2013 was a particularly important year as this was when the Japan Association of Corporate Executives (*keizai dōyūkai*) proposed the use of TOEFL iBT as the national UEE; the ‘Kyoto Appeal’ was announced (see English summary in Mizumoto, 2016; in Japanese, http://www.jasele.jp/wp-content/uploads/kyoto_appeal_2013.pdf); and Otsu et al. (2013) began their counter-assault.

⁵ All quotations from Japanese-language texts are indicated with single quotation marks and are my translations.

⁶ It must be noted that four-skills tests have been much more widely adopted by institutions for alternative admission routes (e.g., the current use of EIKEN, TEAP, and IELTS by universities: <https://search.eiken.or.jp/qualification/exam/>). The present reform seeks to expand the use of four-skills tests to general admissions.

⁷ It is not the purpose of this paper to critique the rationale and empirical support for the approach to language education and assessment outlined in the SCJ proposal. Nor is the purpose to critique each of the authors’ individual views, which are, in fact, quite diverse, while still converging on the principles outlined above. Nevertheless, it should be clear that the numerous claims made in the SCJ are highly controversial (for discussion on the intersection between pedagogical approaches and second language acquisition, see e.g., Ellis & Shintani, 2014). Indeed, all indications are that the position is essentially a hark back to the grammar-translation approach that was widespread in the pre-communicative era.

⁸ As stated previously, the definition of UEE here does not include the Common Test and NCT. Detailed information of test specifications and validation is provided about these centrally developed tests (see Watanabe, 2013; National Centre for University Entrance Exams, <https://www.dnc.ac.jp/kyotsu/index.html>).

References

- Abe, M. (2017). *Shijō saiaku no eigo seisaku – uso-darake no ‘4 ginō’ kanban* [English education in chaos: Confusion and dishonesty in Japanese government policy]. Hitsuji Shobō.
- Abe, M. (2018). Naze supīkingu nyūshi de, supīkingu-ryoku ga ochiru no ka. [Why will speaking ability decrease through use of a speaking entrance exam?] In T. Haebara (Ed.), *Kenshō meisō suru eigo nyūshi — supīkingu dōnyū to minkan itaku —* (pp. 69–88). Iwanami Shoten.
- Allen, D. (2016a). Investigating washback to the learner from the IELTS test in the Japanese tertiary context. *Language Testing in Asia*, 6(7), 1–20. <https://doi.org/10.1186/s40468-016-0030-z>
- Allen, D. (2016b). Japanese cram schools and entrance exam washback. *Asian Journal of Applied Linguistics*, 3(1), 54–67. <https://www.caes.hku.hk/ajal/index.php/ajal/article/download/338/412>

- Allen, D. (2017). Investigating Japanese undergraduates' English language proficiency with IELTS: Predicting factors and washback. *IELTS Partnership Research Papers 2*. IELTS Partners. <https://www.ielts.org/teaching-and-research/research-reports>
- Allen, D. (2020). *Shiken*: Past and future. *Shiken*, 24(1), 22-36. http://teval.jalt.org/sites/teval.jalt.org/files/24_01_22_Allen_Shiken_past_and_future.pdf
- Allen, D. & Nagatomo, D. H. (2019). Investigating the consequential validity of TEAP: Washback to high school learners of English. *Eiken Research Report*. Eiken Foundation of Japan. https://www.eiken.or.jp/center_for_research/pdf/bulletin/vol99/vol_99_21.pdf
- Baba, S. (2019). Dōsureba haisuteikusu tesuto no nozomashī hakyū kōka o mo tarasu koto ga dekiru no ka? Kyōiku shinri-gaku kenkyū kara no teian [How to produce beneficial washback effect by using high-stakes testing? Proposal from educational psychology]. *JLTA Journal*, 22, 44–64. https://doi.org/10.20622/jltajournal.22.0_44
- Bunch, M. B. (2012). *Aligning curriculum, instruction, and assessment*. Measurement Inc. <https://measurementinc.com/white-papers>
- Brown, J. D. (1996, February 5). Japanese entrance exams: A measurement problem? *The Daily Yomiuri* (Educational Supplement), 15.
- Brown, J. D., & Yamashita, S. O. (1995a). English language entrance examinations at Japanese universities: What do we know about them? *JALT Journal*, 17(1), 7–30. <https://jalt-publications.org/sites/default/files/pdf-article/jj-17.1-art1.pdf>
- Brown, J. D. & Yamashita, S. O. (1995b). English language entrance examinations at Japanese universities: 1993 and 1994. In J. D. Brown and S. O. Yamashita (Eds.) *Language Testing in Japan*. Japanese Association for Language Teaching.
- Brown, J. D., & Yamashita, S. O. (1995c). The authors respond to O'Sullivan's letter to JALT Journal: Out of criticism comes knowledge. *JALT Journal*, 17(2), 257-260. <https://jalt-publications.org/sites/default/files/pdf-article/jj-17.2-art8.pdf>
- Buck, G. (1989). Written tests of pronunciation: Do they work? *ELT Journal*, 43(1), 50-56. <https://doi.org/10.1093/elt/43.1.50>
- Butler, Y., & Iino, M. (2005). Current Japanese reforms in English language education: The 2003 "action plan". *Language Policy*, 4(1), 25-45. <https://doi.org/10.1007/s10993-004-6563-5>
- Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge University Press. <https://rm.coe.int/16802fc1bf>
- Council of Europe. (2018). *Common European Framework of Reference for Languages: Learning, teaching, assessment (Companion volume with new descriptors)*. <https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989>
- Ellis, R., & Shintani, N. (2014). *Exploring language pedagogy through second language acquisition research*. Routledge.
- Fujiwara, Y. (2018). Daigaku eigo nyūshi de nani o hakarubeki ka [What to measure in college English entrance exams]. *JACET Chūbu shibu kiyō*, 16, 1–32. https://researchmap.jp/yasuhiro008/published_papers/18506664

- Green, A. (2014). *The Test of English for Academic Purposes (TEAP) impact study: Report 1 - preliminary questionnaires to Japanese high school students and teachers*. Eiken Foundation of Japan. https://www.eiken.or.jp/teap/group/pdf/teap_washback_study.pdf
- Haebara, T. (Ed.) (2018). *Kenshō — meisō suru eigo nyūshi — supīkingu dōnyū to minkan itaku —* [Inspection of stray English entrance exams – Introduction of private company speaking exams]. Iwanami Shoten.
- Hasegawa, S., Chujo, K., & Nishigaki, C. (2006). Daigaku nyūshi eigomondai goi no gaido to yūyō-sei no jidai-teki henka [A chronological study of the level of difficulty and the usability of the English vocabulary used in university entrance examinations]. *JALT Journal*, 28(2), 115–134. <https://doi.org/10.37546/JALTJJ28.2-1>
- Henrichsen, L. E. (1989). *Diffusion of innovations in English language teaching: The ELEC effort in Japan, 1956–1968*. Greenwood Press.
- Japan Language Testing Association (JLTA). (2017). *Proposal for handling English testing within the 'Prospective university entrance scholastic abilities evaluation test [provisional name]'*. http://jlta2016.sakura.ne.jp/?page_id=865
- Kikuchi, K. (2006). Revisiting English entrance examinations at Japanese universities after a decade. *JALT Journal*, 27(1), 77–96. <https://doi.org/10.37546/JALTJJ28.1-5>
- Koizumi, R. (2018). *Eigo 4 ginō tesuto no erabikata to tsukaikata: Datōsei no kanten kara* [How to choose and use four-skills tests: From the viewpoint of validity]. ALC.
- Kuramoto, N. & Koizumi, R. (2018). Current issues in large-scale educational assessment in Japan: Focus on national assessment of academic ability and university entrance examinations, *Assessment in Education: Principles, Policy & Practice*, 25(4), 415-433, <https://doi.org/10.1080/0969594X.2016.12256677>
- The Mainichi. (2019, October 28). *Japan minister apologizes for comments downplaying inequality among university test takers*. <https://mainichi.jp/english/articles/20191028/p2a/00m/0na/013000c>
- MEXT. (2016). *Kōdai setsuzoku shisutemu kaikaku kaigi: Saishū hōkoku* [The final announcement of reports on discussions regarding the improvement of the upper secondary school-university articulation]. https://www.mext.go.jp/b_menu/shingi/chousa/shougai/033/toushin/1369233.htm
- MEXT. (2017a). *Gaikokugokatsudō gaikoku-go-hen: Shōgakkō gakushū shidō yōryō (Heisei 29-nen kokuji) kaisetsu*. [Foreign language activities / foreign languages: Explanation of elementary school curriculum guidelines (2017 notification)]. https://www.mext.go.jp/component/a_menu/education/micro_detail/_icsFiles/afieldfile/2019/03/18/1387017_011.pdf
- MEXT. (2017b). *Gaikoku-go-hen: Chūgakkō gakushū shidō yōryō (Heisei 29-nen kokuji) kaisetsu* [Foreign language edition: Explanation of junior high school curriculum guidelines (2017 notification)]. https://www.mext.go.jp/component/a_menu/education/micro_detail/_icsFiles/afieldfile/2019/03/18/1387018_010.pdf
- MEXT. (2018a). *Gaikoku-go-hen eigo-hen: Kōtōgakkō gakushū shidō yōryō (Heisei 30-nen kokuji) kaisetsu* [Foreign languages (English): Explanation of high school curriculum guidelines (2018 Notification)].

https://www.mext.go.jp/component/a_menu/education/micro_detail/_icsFiles/afieldfile/2019/03/28/1407073_09_1_1.pdf

MEXT. (2018b). *Kaku shikaku kentei shiken to CEFR to no taishō-hyō* [Comparison table between each qualification / certification test and CEFR].

https://www.mext.go.jp/b_menu/houdou/30/03/_icsFiles/afieldfile/2019/01/15/1402610_1.pdf

MEXT. (2019). *Reiwa 3-nendo daigaku nyūgaku-sha senbatsu ni kakaru daigaku nyūshi eigo seiseki teikyō shisutemu un'ei taikō no haishi ni tsuite (tsūchi)* [Regarding the abolition of the university entrance examination English grade provision system management outline for the selection of university enrollees in the 3rd year of Reiwa (Notice)]

https://www.mext.go.jp/a_menu/koutou/koudai/detail/1420229.htm

MEXT. (2020a). *Daigaku nyūgaku-sha senbatsu kanren kiso shiryōshū* [Collection of basic materials related to selection of university enrollees]. https://www.mext.go.jp/content/20200318-mxt_daigakuc02-000005103_8.pdf

MEXT. (2020b). *Daigaku nyūshi ni kansuru web iken boshū ni tsuite* [About soliciting web opinions about university entrance exams]. https://www.mext.go.jp/content/20200929-mxt_daigakuc02-000009870_3.pdf

MEXT. (2020c). *Daigaku nyūshi no arikata ni kansuru kentōkai* [Examination meeting regarding the ideal way of university entrance examination].

https://www.mext.go.jp/b_menu/shingi/chousa/koutou/103/index.htm

Mizumoto, A. (2016). Introducing Kyoto Appeal: Issues in and implications of using four-skills proficiency tests as entrance examinations in Japan. *British Council new directions in language assessment: JASELE journal special edition*, 59-68. <http://hdl.handle.net/10112/13018>

Moore, Y. (2015). An evaluation of English writing assessment in Japanese university entrance examinations. *Writing & Pedagogy*, 7(2), 233–260. <https://doi.org/10.1558/wap.v7i2-3.26227>

Munby, J. (1978). *Communicative syllabus design*. Cambridge.

Nagatomo, D. H. & Allen, D. (2019). Investing in their futures: Highly-motivated students' perceptions of TEAP and university entrance exam. *The Language Teacher*, 45(5), 3-7. <https://jalt-publications.org/sites/default/files/pdf-article/43.5-tlt-art1.pdf>

O'Sullivan, B. (1995). A reaction to Brown and Yamashita "English language entrance exams at Japanese universities: What do we know about them?" *JALT Journal*, 17(2), 255-257. <https://jalt-publications.org/sites/default/files/pdf-article/jj-17.2-art7.pdf>

O'Sullivan, B. (2020). *The Comprehensive Learning System*. British Council White Papers on English Language Policy & Education. British Council.

Otsu, Y. (2013). *Eigo kyōiku seisaku wa naze machigau no ka: Ninchi kagaku gakushū kagaku no shiten kara* [Why is the English education policy wrong? From the perspective of cognitive science and learning science]. In Y. Otsu, H. Erikawa, Y. Saito, & K. Torikai (Eds.), *Eigo kyōiku, semari kuru hatan* (pp. 51-72). Hitsuji Shobō.

Otsu, Y., Erikawa, H., Saito, Y., & Torikai, K. (Eds.) (2013). *Eigo kyōiku semari kuru hatan* [English education in danger]. Hitsuji Shobō.

Sasaki, M. (2008). The 150-year history of English language assessment in Japanese education. *Language Testing*, 25(1), 63–83. <https://doi.org/10.1177/0265532207083745>

- Sato, T. (2018). The impact of the Test of English for Academic Purposes (TEAP) on Japanese students' English learning. *JACET Journal*, 62, 89–107. https://doi.org/10.32234/jacetjournal.62.0_89
- Sato, T. (2019). An investigation of factors involved in Japanese students' English learning behavior during test preparation. *Papers in Language Testing and Assessment*, 8(1), 69-95. http://www.altanz.org/uploads/5/9/0/8/5908292/8_1_s4_sato.pdf
- Science Council of Japan, Language and Literature Committee (2020). *Daigaku nyūshi ni okeru eigo shiken no arikata ni tsuite no teigen* [Recommendations for the English test for university entrance exam]. <http://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-24-t292-6.pdf>
- Torikai, K. (2013). *Eigo komyunikēshon nōryoku wa hakareru ka* [Can English communication skills be measured?] In Y. Otsu et al. (Eds.), *Eigo kyōiku, semari kuru hatan* (pp. 83-116). Hitsuji Shobō.
- Torikai, K. (2018). *Eigo kyōiku no kiki* [Crisis of English education]. Chikuma Shobō.
- Watanabe, Y. (1996). Does grammar translation come from the entrance examination? Preliminary findings from classroom-based research. *Language Testing*, 13(3), 318-333. <https://doi.org/10.1177/026553229601300306>
- Watanabe, Y. (1997). *The washback effects of the Japanese university entrance examinations of English - classroom-based research* [Unpublished doctoral dissertation]. Lancaster University.
- Watanabe, Y. (2004). Teacher factors mediating washback. In L. Cheng, Y. Watanabe & A. Curtis (Eds.), *Context and method in washback research: The influence of language testing on teaching and learning*, (pp. 129–146). Lawrence Erlbaum.
- Watanabe, Y. (2013). The National Center Test for University Admissions. *Language Testing*, 30, 565–573. <https://doi.org/10.1177/0265532213483095>
- Weir, C. J. (2005). Limitations of the Common European Framework for developing comparable examinations and tests. *Language Testing*, 2(3), 281–300. <https://doi.org/10.1191/0265532205lt309oa>
- Yanagawa, K. (2012). *A partial validation of the contextual validity of the Centre listening test in Japan* [Unpublished doctoral dissertation]. University of Bedfordshire. <https://uobrep.openrepository.com/handle/10547/267493>
- Yoshida, K. (1996, January 15). Language testing in Japan: A cultural problem? *The Daily Yomiuri* (Educational Supplement), 15.