Usability of a speaking assessment portal for Japanese teachers of English

Rie Koizumi, Makoto Fukazawa, and Chihiro Inoue
koizumi.rie.ge@u.tsukuba.ac.jp
fukazawa@edu.u-ryukyu.ac.jp
chihiro.inoue@beds.ac.uk
University of Tsukuba
University of the Ryukyus
University of Bedfordshire

Abstract

Against a backdrop of insufficient training for pre-service and in-service teachers, as well as limited access to materials and resources related to speaking assessment (SA), this study reports on the development and usability of an SA Portal, drawing upon the perceptions of teachers who used the website. The Portal is intended for Japanese senior high school teachers of English as a way to equip them with a wider range of relevant resources. It includes tips for conducting speaking tests; SA examples and explanations, including tasks, rubrics, and videos; and useful websites and resource. There were two phases in this preliminary usability study, and teacher perceptions were collected in each phase. We found that teachers received the content of the Portal positively. Teachers also provided numerous points for improvement from micro and macro levels. Most of these suggestions have been implemented in the Portal, while the remaining ones will be considered in the future. The practical implications of the Portal itself and the use of feedback from its users are also outlined. Specifically, soliciting input from users with diverse backgrounds, employing various open-ended questions, and allowing sufficient time for multiple revisions can lead to valuable feedback that contributes to effective improvements.

Keywords: second language speaking assessment, online resources, teacher training, rubric, teacher perception

Speaking assessment (SA) in classrooms is an indispensable element of language education (e.g., Poehner & Inbar-Lourie, 2020). It can be used for formative and summative purposes, helping both teachers and students understand the students' learning status, strengths, weaknesses, and other features. In this context of classroom-based speaking assessment, teachers act as the primary test developers, administrators, raters, and providers of feedback. The Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT) strongly encourages second language (L2) English teachers in Japan to use performance-based SA in classrooms to assess knowledge and skills; thinking, judgment, and expression; and a proactive attitude towards learning (i.e., linguistic accuracy, content appropriateness, and willingness to communicate; National Institute for Educational Policy Research, 2023). However, the frequency and quality of SA need to be improved (Koizumi, 2022a, 2022b; Tando, 2023; see Kaneko, 2019, for teachers' voices for this issue). To enhance the quantity and quality of SA in Japan, two issues must be addressed with urgency (e.g., Koizumi, 2022b). First, there are insufficient opportunities for both pre-service and in-service Japanese teachers to receive proper training. Second, there is a lack of materials and resources related to SA, especially those freely available online in Japanese. As a result, teachers often lack opportunities to learn how to select appropriate SA formats and rubrics from various options and how to use them consistently and formatively (e.g., Koizumi, 2022b). Increasing the availability of resources for teachers would help enhance teachers' L2 assessment literacy by incorporating these resources into teacher training programs or help build consensus among teachers within and across their respective schools.

To address these issues, we developed a Speaking Assessment (SA) Portal (the Portal, hereafter), available online for Japanese teachers of English, particularly in senior high schools. It is expected that teachers will use the Portal for teacher training, independent self-study, and teacher meetings at local and regional

levels, and that they can enhance their L2 assessment literacy to implement SA effectively in the classroom. While reporting on the development of the Portal, we also present a preliminary usability study based on teachers' perceptions of the Portal, using two datasets.

Literature Review

Multiple online resources are available for L2 teachers and those interested in learning about language assessment and SA. For example, *Language Testing Resources Website* (Fulcher, 2024) has been prominent in disseminating essential knowledge, including useful videos, explanations, and discussion topics on language assessment in English. Assessment & Evaluation Language Resource Center, Georgetown University (2024) provides a summary of resources for teachers to learn about language assessment in English. British Council (2024b) provides a practical glossary and videos on language assessment in English. They also provide helpful videos in Japanese with a focus on the assessment of four skills (British Council, 2024a). British Council (n.d.) also hosts useful teacher training kits in English. The Japan Language Testing Association (n.d.) also hosts various functional workshop videos and online tutorials, primarily in Japanese. Another existing resource is Tools to Enhance Assessment Literacy (TEAL, 2024c), which focuses on broad aspects of language assessment, particularly in the context of teaching additional languages (e.g., Vietnamese and Tagalog) in Australia.

Many English testing resources are open access (e.g., *Language Testing Resources Website*; Fulcher, 2024), which help teachers acquire the fundamental knowledge of language assessment. Among these, TEAL (2024c) is considered the most beneficial in comprehensively providing not only guidelines on how to implement SA but also numerous concrete examples (TEAL, 2024a). These include 21 SA tasks (e.g., "Role play: Giving advice to a friend"); and for each task, a rubric; three to seven learners' videos; and a commentary for each video.

However, the usability of TEAL's (2024a) task and other examples are limited for Japanese teachers, primarily because MEXT recommends using a specific rubric format in classroom SA, which differs from that in TEAL (MEXT, 2022; National Institute for Educational Policy Research, 2023). Moreover, speech samples in the videos are not always similar to those typically produced by Japanese learners. Building on the need for SA resources that specifically cater to Japanese teacher audiences, we created the online Portal and conducted a usability study by obtaining feedback from teachers to refine the quality of the Portal.

The research questions (RQs) are as follows:

1. After using the Portal, how do Japanese teachers of English perceive its usefulness?

2. What do these teachers indicate as areas for improvement?

Method

Speaking Assessment (SA) Portal

Using TEAL as a model, we developed the Portal (https://sites.google.com/view/speaking-assessment/) including the principles and practices of developing and conducting SA (i.e., tasks, rubrics, speech samples, explanations of how they are scored), and resources for further learning. Figure 1 shows the top page of the Portal, which is freely accessible to anyone. As shown in Table 1, the Portal includes sections such as "Tips for conducting speaking tests," "SA examples and explanations," "Useful websites and resources," and "Frequently Asked Questions" on test development, administration, scoring, feedback, and other SA matters.

Figure 1
Top Page of the Speaking Assessment Portal

Table 1
Structure of Speaking Assessment (SA) Portal (as of September 2024)

Section	Content
1. Tips for conducting speaking tests [J]	How to develop speaking tests How to administer speaking tests How to score elicited spoken performance Types of feedback to provide
2. SA examples and explanations [J]	14 tasks (including both monologues and dialogues): role plays with a teacher, teacher-led interviews, oral interaction in pairs, and short speeches with questions and answers among paired students Each task includes a task description, a rubric, a worksheet, and six to 10 videos of learner speech samples per task. 120 videos in total[a]. Each video is edited to blur parts that could reveal personal information. Each video is accompanied by scores based on the rubric, a transcription, and a rationale for the given scores.
3. Useful websites and resources (to direct users to resources for further learning) [J]	Videos, scoring criteria, and sample scores from various speaking tests and assessment practices, with each video classified by proficiency levels (e.g., Graded Examinations in Spoken English [GESE] and Integrated Skills in English by Trinity College London)

	Resources for developing one's own speaking tests (e.g., task examples in various tests, analytical tools, Interactional Competence checklist [full and brief versions of Nakatsuhara et al., 2018, translated into Japanese], and materials created by municipal boards of education[b])
	Resources for updating L2 assessment literacy related to SA (e.g., free online courses and resources such as Instructional Topics in Educational Measurement Series, and Language Assessment in the Classroom])
	Resources for learning about automated scoring[b]
4. Frequently Asked Questions (FAQ)[b] [J]	Answering questions regarding test development, administration, scoring, giving feedback, and other SA matters
5. Research meetings[b] [J]	Language Learning Assessment Research Meetings
6. Digest of SA in Japan[b] [E]	Videos demonstrating and explaining optimal scoring practices for SA
7. Project members [J]	Introduction of the members who contributed to the development of the Portal

Note. [ ] = Language used; J = Japanese; E = English. [a] In Phase 1, the Portal had five tasks and approximately five videos per task. [b] Uploaded after Phase 1.

The Portal mainly differs from TEAL in that it specifically focuses on SA tasks and rubrics that adhere to MEXT guidelines, and speech sample videos with first language (L1) Japanese speakers learning English as an L2. The Portal is tailored to the Japanese context of learning English as a foreign language. It utilizes the learners' L1 and it addresses narrower ranges of English proficiency levels and learner profiles. Tasks vary from role plays with a teacher, teacher-led interviews, oral interaction in pairs, and short speeches with questions and answers among the paired students. Role-play tasks with teachers were originally developed as part of the CEFR-J project and linked to CEFR-J levels (see Tono & Negishi, 2020, for task development; see Koizumi, 2022a; Tono, 2022, for actual tasks).

Usability Study

The project to improve the website consisted of two phases, each involving Japanese teachers of English. In Phase 1 (April 2020 to March 2022), the Portal was planned and created by the authors, and tested by teachers through online questionnaires. The teachers' feedback was used for substantial revisions. We also presented the Portal and its development at a conference where we received additional feedback from the audience. In Phase 2 (April 2022 to November 2023), we further revised the Portal and sought feedback from another group of teachers. All instructions to the participants and feedback from them were provided in Japanese. All direct citations were translated from Japanese to English by the first author.

Participants and Procedures in Phase 1

We recruited six Japanese teachers of English with more than 10 years of teaching experience to participate in this study (Teachers A to F in Table 2). We intended to diversify the study participants to obtain feedback from various perspectives. An honorarium was provided, except for one participant who declined to receive it.

Table 2
Summary Statistics of Persons

Phase	Participant	Background (Approximate time spent in Phase 1)
1	Teacher A	Taught English in a senior high school (2 hours)
	Teacher B	Taught English in a senior high school (1.5 hours)
	Teacher C	Worked for a prefectural education in-service training center, responsible for training teachers; previously taught English in a senior high school (3 hours)
	Teacher D	Taught English language teaching at a university education department; previously taught English in a junior high school (3 hours)
	Teacher E	Retired from a university, specialized in language assessment; previously taught English in a senior high school (2 hours)
	Teacher F	Worked for a private company after teaching English in a junior high school (5.5 hours)
2	Teacher G	Taught English in an elementary school
	Teacher H	Taught English in a junior high school
	Teacher I	Taught English in a senior high school

The six teachers in Phase 1 answered Questionnaires 1 and 2, which included closed- and open-ended questions (see Appendices A to D for questions in Japanese and English). First, they watched a video with instructions regarding what they were going to do. They were informed that the main target users were senior high school English teachers, although the Portal may also provide useful information to English teachers at other types of schools. They were requested to answer as if they were teachers who administered speaking tests to students and to examine the usefulness, appropriateness, and ease of content to improve the quantity and quality of the Portal. Second, they were asked to spend 30 minutes browsing through the overall Portal and wrote their opinions and suggestions for improvement in Questionnaire 1. Third, they were requested to spend approximately 1.5 hours reading "SA examples and explanations" and answering the questions in Questionnaire 2. The participants spent approximately 1.5 to 5.5 hours reading the Portal and answering all the questions.

Participants and Procedures in Phase 2

We solicited additional feedback from three teachers (Teachers G to I in Table 2). Although we intended to create the Portal to primarily cater to senior high school teachers, we also included elementary and junior high school teachers to explore the potential usefulness and challenges of expanding our focus. We asked the teachers to provide an overall impression of the Portal focusing on its useful aspects and those that need to be improved. They presented their perspectives in a PowerPoint file and discussed their opinions at an online research meeting. An honorarium was provided afterward.

Analysis in Phases 1 and 2

Responses to the closed questions in Phase 1 of Questionnaires 1 and 2 were tallied. Verbal feedback in the open-ended format in Phases 1 and 2 was analyzed thematically.

Results and Discussion

Overall Perceptions of the Portal (RQ1)

We generally received positive responses from teachers in both Phases 1 and 2. Therefore, we report the results from both phases together in this section. As shown in Table 3, most teachers in Phase 1 found some of the content interesting and appropriate. "SA examples and explanations" was considered the most interesting by five teachers, followed by "Tips for conducting speaking tests" and "Useful websites and resources," each selected by four teachers. Similarly, most teachers found "SA examples and explanations" and "Useful websites and resources" the most appropriate (five teachers), followed by "Tips for conducting speaking tests" (three teachers).

Table 3

Number of Teachers Who Found the Portal Content Interesting and Appropriate in Phase 1

	Tips for conducting speaking tests	SA examples and explanations	Useful websites and resources	Project members
Interesting[a]	4	5	4	1
Appropriate[b]	3	5	5	2

Note. n = 6.[a] Based on Questionnaire 1, Item 2.[b] Positive comments were tallied based on responses in Questionnaire 1, Items 3 to 5 and Questionnaire 2, Item 3. See Appendices A to D for actual items.

Here is an example comment from a teacher, regarding SA examples and explanations:

> The videos provided cover various task formats to a certain degree. After understanding the key points about SA through the five tasks, teachers should be able to adapt the format to other tasks with different topics and situations. All the scoring procedures—specifically, how teachers use the rubric to score— are easy to understand with the provided transcriptions and rationales for the scores. The edited conditions of the videos were useful. Although parts of the videos were blurred, the atmosphere during the interaction was easily understandable. (Teacher D)

Teachers C and E mentioned that "Useful websites and resources" help teachers understand how to conduct interviews by providing level-specific videos and scoring rubrics from speaking tests in other countries. Teacher C also noted that this type of online resource is much more useful and effective than paper-based booklets.

In Phase 2, three teachers provided positive feedback on the topics and content of the Portal as found below:

> Essential points, such as how to conduct speaking tests, are summarized on the Portal. Many sample videos help teachers visualize the tests. Teachers can also learn how to score the tests by reading the transcriptions. (Teacher H)

> By watching video explanations of scores in each video, teachers can learn how to set a scene and situation in a speaking test and how teachers can respond and ask questions according to students' proficiency levels and their reactions. By reading the guidelines on the Portal, it was easy to see how to administer and score speaking tests. For example, normally it is difficult for teachers to find time to explain SA tests and conduct them during class time. However, the explanations on the Portal were useful for creating speaking tests with a balanced focus on validity, reliability, and practicality. Even for current teachers, the Portal can supplement insufficient in-service teacher training. (Teacher G)

The explanations on development, administration, scoring, and giving feedback, as well as example rubrics and CEFR information, could be helpful when teachers attempt to relate teaching with assessment in the classroom and create tasks in their own contexts. The Portal helps teachers grasp the gist of SA before reading Koizumi (2022a) in detail. (Teacher I)

Areas for Improvement That We Addressed and Will Address (RQ2)

While teachers had a positive view of the overall design and content of the Portal, they also offered suggestions for further modifications during Phases 1 and 2. Below, we summarize these suggestions according to categories, rather than by phases, using direct quotations from the teachers. We also explain how we addressed these suggestions under "Solution." Areas for improvement that we have not yet addressed are summarized under "Future plan."

Overall design (from Phase 1)

- Making explanations easy to read

    - I understand that writers try to use simple language, but descriptions are sometimes difficult for high school teachers to understand. (Teacher D)

    - There are many words on the page. There should be a blank line between (1) and (2). (Teacher F)

        - Solution: We decreased technical terms, simplified the language, and added blank lines between the items. We also provided a subsection "Further reading" for those interested.

- Ensuring consistent use of terms:

    - Terms are not always consistent, such as *raters* and *scorers* (i.e., *hyokasha*, *saitensya*). (Teacher F)

        - Solution: We revised the site to use the term *saitensya* and other terms consistently.

- Using a clear and consistent layout:

    - The layout must be consistent to ensure a unified atmosphere. (Teachers D and F)

    - Font sizes should be larger and consistent across sections. (Teachers C and D)

    - Color and fonts should be used to show the highlights and make reading easy. (Teacher D)

    - The layout in 'SA examples and explanations' with bar buttons to show URLs is much clearer than 'Tips for conducting speaking tests.' The addition of illustrations is helpful. (Teacher D)

    - Regarding 'SA examples and explanations,' having a bar button in blue for Example 1 would enhance handling ease. Having red letters corresponding to the criteria, rather than simple black letters, makes it easier to understand. (Teacher D)

    - Illustrations in 'Useful websites and resources' are too large (Teacher F)

        - Solution: We modified the layout and used consistent colors and fonts across sections and items.

- Avoiding mechanical errors:

- ▪ While I understand that the Portal is under construction, there are many noticeable typos. (Teacher A)

  - ▪ Solution: We reviewed and corrected all content.

- • Providing access to videos:

  - ▪ YouTube videos cannot be viewed in teacher rooms in certain prefectures. Although some teachers may be able to watch them on tablets, many schools only have tablets for students. Some teachers found it difficult to watch the YouTube videos. (Teacher B)

    - ▪ Solution: We added an explanation to the FAQ section that videos can be obtained by contacting us.

  - ▪ I could not view the videos in French, in the 'Useful websites and resources.' section (Teacher C)

    - ▪ Solution: We corrected the URL.

Suggestions for Tips for conducting speaking tests in the Portal (from Phase 1)

- • Providing a brief introduction of speaking tests:

  - ▪ A section is needed to briefly explain what performance tests look like. Including various test formats and patterns, such as student-teacher and student-student conversations, would help young teachers transition to other more detailed pages. (Teacher D)

    - ▪ Solution: We provided this information.

- • Explaining how to modify tasks and rubrics:

  - ▪ (Scores based on a rubric are provided for each video performance.) One of the biggest concerns is that teachers at lower-level schools may consider speaking tests unmeaningful when they see a video where all scores are cs [out of a, b, and c, with c being the lowest score]. They may think that all of their students would receive similar scores. Providing an explanation of how to modify the rubric and guidelines to conduct speaking tests according to their context would be helpful. (Teacher C)

    - ▪ Solution: We added an explanation to the FAQ section of the Portal.

- • Explaining how to conduct speaking tests:

  - ▪ It is ideal to have case scenarios for conducting speaking tests for approximately 40 students per class, with five classes in one school per year. Conducting speaking tests to measure interaction (dialogues) and presentation (monologues) efficiently and fairly would be helpful. (Teacher C)[1]

  - ▪ Considerations for teachers to conduct a role play with a student would be helpful in understanding basics (e.g., how to take a neutral stance to avoid any effects of teachers on student performance; how to create a supportive atmosphere in which students can speak well). (Teacher F)

    - ▪ Solution: We added explanations to the FAQ section.

- • Explaining how to score speaking and reach final scores:

- It is easy to judge the number of sentences to score, but it is difficult to judge which is better: two short sentences without conjunctions or a long sentence with conjunctions. The explanation would be helpful in this regard. (Teacher D)

    - Solution: We added an explanation to the FAQ section.

- An explanation and example video performances on how to finalize scores when they differ across teachers would be helpful. (Teacher D)

    - Solution: We added an explanation in the subsection of Scoring speaking tests.

    - Future plan: We will consider including examples in the future.

- I would like to know how and where teachers diverge in scoring, even after discussing the rubric beforehand. (Teacher A)

    - Future plan: We will address this in the future.

- Explaining matters related to MEXT evaluation guidelines:

    - Having only three levels for all evaluations was unreasonable. Therefore, a fine-grained evaluation would be more appropriate. (Teacher E)

    - Evaluating the willingness to communicate is difficult. Students usually try to speak during a test, so they will eventually receive Score b [out of a, b, and c, with c being the lowest score]. (Teacher E)

        - Solution: Because the Portal follows MEXT's evaluation guidelines, we explained it as is. We also noted this on the top page of 'SA examples and explanations' and added it to the FAQ section.

- Explaining how to create videos or recordings:

    - An explanation would be helpful for technical matters important in developing and administering speaking tests, such as how to videotape and record performance. (Teacher E)

        - Solution: We added an explanation to the FAQ section.

- Providing concrete examples:

    - Examples of feedback explanations are needed, such as samples of feedback on sheets and a video on giving oral feedback, and examples and explanations of score reports, which would help teachers understand the image of this activity. (Teacher F)

        - Solution: We added the explanation to the section.

- Providing resources for further learning:

    - The Portal says that feedback should include not only the current speaking ability, but also how to improve it. Any website that helps increase speaking ability and is accessible to students during self-study would be helpful. (Teacher F)

        - Solution: We added an explanation to the FAQ section.

- Providing downloadable materials:

- ▪ I would like to have rubrics and worksheets (also feedback sheets and reflection worksheets) downloadable in Excel and Word, which I can modify according to my context. This will save time in developing them myself. (Teacher A)

- ▪ It is important to score while watching the videos. Providing a scoring worksheet would help teachers individually and, in a group, allow them to write scores and rationales. (Teacher C)

    - ▪ Solution: We uploaded the files to the Portal.

Suggestions for SA examples and explanations in the Portal (from Phase 1)

- Providing visual aids:
    - ▪ Along with verbal explanations, a flowchart of explanations and a video explaining SA procedures would be helpful in catering to teachers' individual needs and preferences. (Teacher D)
        - ▪ Future plan: This should be addressed in future revision.

- Providing additional examples:
    - ▪ Having examples from both analytic and holistic rubrics would be helpful. (Teacher E)
        - ▪ Solution: We included analytic examples in the format provided by MEXT as part of the test specification examples. We will consider including holistic rubrics in the future, but analytic rubrics would generally fit the teaching context in Japan.

- Adding an interactive mode:
    - ▪ It may be useful to have a section in which teachers can input their scores and check whether their scores are correct as part of the practice. This gamification may enhance teachers' interest. (Teacher C)
        - ▪ Solution: We considered this option but decided not to include it because such a function might imply that there are absolutely correct answers in scoring performance, which is not our intention. Since scoring rubrics should be tailored to students and various classroom contexts, our focus is on presenting the principles, possible options, and examples of SA practice.

- Changing the order of items:
    - ▪ Reading a rubric before watching a video is intuitively easy to understand. I do this during self-training and discuss the criteria with my colleagues. (Teacher A)
        - ▪ Solution: We changed the order to make the website more user-friendly for teacher training.

- Modifying the length of task explanation in the video:
    - ▪ Having a long time to read the task description is unnecessary. (Teacher E)
        - ▪ Solution: Originally, each slide was shown for 14 seconds, but we shortened it to 7 seconds in the videos.

- Improving video quality:

- ▪ Some videos were difficult to hear due to the recording quality and the students' voice volumes. (Teacher F)

    - ▪ Solution: We added the explanation to the FAQ section.[2]

Suggestions for SA examples and explanations in the Portal (from Phase 2)

- • Providing more fine-grained task specifications:

    - ▪ Task descriptions are broadly written to make tasks generalizable across contexts, although some tasks have specific conditions. According to the Course of Study or the MEXT curriculum guidelines, setting a clear purpose (why you need to do this), scene (in what scene do you talk to), and situation (to whom you are talking) in which students need to communicate in English is important. It may be necessary to emphasize the need for teachers to set concrete and detailed purposes, scenes, and situations while considering class activities and observing students' reactions, NOT using the same task and the rubric from the Portal. (Teacher I)

        - ▪ Solution: We added an explanation to SA examples and explanations.[3]

- • Providing a wider range of tasks, rubrics, topics, and examples:

    - ▪ The current Portal has more interactive tasks, which is nice. However, more monologic tasks would be helpful, as more teachers conduct monologic speaking tests. Furthermore, junior and senior high school students are expected to work on both daily and social topics. The Portal has more daily or familiar topics, and more examples of social topics, such as environmental issues, racial discrimination, and technology, which appear in textbooks, would increase its value. (Teacher I)

    - ▪ While I understand that the Portal is mainly intended for senior high school teachers, the SA format and rubric examples are beyond the level of elementary school students. If easier examples are provided, this will be more helpful. Furthermore, more tasks would allow teachers to understand task variation, such as using class interaction as part of assessment and evaluating recordings submitted by students. (Teacher G)

        - ▪ Future plan: We will consider including such tasks, rubrics, and examples in the future to cater to various needs.

- • Providing additional examples and clarifications for the rubric:

    - ▪ Regarding 'Willingness to communicate' in the rubric, it is difficult to understand what behaviors and utterances are measured in evaluating students 'trying to communicate to the partner(s),' although this may depend on each school's context. Regarding "Content appropriateness" in the rubric, questions arise as to (a) whether utterances need to be sentences, not fragments, and (b) which is evaluated more highly: detailed utterances with grammatical errors OR brief utterances with correct grammar. More detailed examples of the rubric would also be helpful. (Teacher H)

        - ▪ Future plan: We will include such examples and clear explanations, although each teacher or school needs to plan practices themselves eventually.

- • Explaining how to select representative videos:

    - ▪ It is great that the Portal contains 120 videos. However, it is difficult to watch all of them. It might be helpful to show a selection of a few tasks first or to display only the first video,

with the second one appearing after watching the first. This could make it easier for busy teachers to navigate. (Teacher G)

- Solution: We added an explanation to the FAQ section on how the videos are categorized and how a single video can be selected for viewing.

- Future plan: We will further consider creating a suggested entry point.

Suggestions for Useful websites and resources in the Portal (from Phase 1)

- Explaining technical terms:
  - I hope to read more explanations in 'Useful websites and resources' on, for example, what GESE is, and what it does. (Teacher D)
    - Solution: We included more information in a concise manner for teachers.

- Providing visual aids:
  - A concise table of the CEFR levels at the top of the page would be helpful. Teachers would like to examine the relationship between such levels and high school students' first- to third-year levels. (Teacher D)
    - Solution: We added a table along with Eiken grade information.

- Providing additional materials:
  - Practical, leading-edge examples from across Japan would be helpful. (Teacher D)
    - Solution: We included a summary of good practices and useful resources from the websites of municipal boards of education. We also added information on cutting-edge research such as automated scoring.

Suggestions for publicizing the Portal (from Phase 1)

- Taking strategic measures to publicize the Portal:
  - Taking strategic measures to publicize the Portal would attract more visitors. An example is asking municipal boards of education and educational centers across Japan to promote the Portal and actively use its contents in teaching training. Only a limited number of teachers read the monthly *English Teachers' Magazine* (by Taishukan Publishing). Annual training for first-, fifth-, and tenth-year teachers would be particularly beneficial. (Teacher C)
    - Future plan: Contacting university teachers who teach in programs that offer teaching certification courses may also be helpful. Therefore, these measures should be considered in the future. Some teachers already found the Portal and contacted us or reported using it, so we should also check for missing groups to be contacted.

As seen above, the teachers' comments focused on both micro and macro levels. The micro-level feedback included suggestions on visual design and the use of simple language to make the resources more user-friendly and enhance readers' understanding. The macro-level feedback involved recommendations for adding more value to the Portal, such as providing explanations on unexpanded topics and increasing awareness among the intended readers. Thanks to the productive feedback from teachers in Phases 1 and 2, we identified additional areas for improvement, detailed as follows:

- Adding visual aids:
  - We can include a flowchart that helps teachers select an appropriate task example and decide on their test specifications by choosing the ability to be measured, the test format, and/or test requirements.
  - We can hide scores and explanations when teachers watch the videos, revealing them by clicking on a bar. This would allow teachers to focus on watching the video and scoring them by themselves.
- Adding various examples:
  - We can include various rubrics and score examples based on a single video (e.g., providing cases of strict and lenient rubrics and scoring decisions).
  - We can include videos showing how students develop their speaking abilities over time, allowing teachers to intuitively understand the students' longitudinal progress (see Tamura, 2022, for such videos).
  - We can include content to help teachers to understand how students' spoken utterances differ depending on test formats during the same period.
- Adding a test task bank:
  - We can include a bank of test tasks and a rubric (i.e., test task bank), which is a concept similar to the Task-Based Language Teaching [TBLT] Language Learning Task Bank (Indiana University, 2024). However, a test task bank would differ from the TBLT Task Bank by providing information on test difficulty and other measurement details (see Koizumi, 2022b).
- Adding an interactive platform:
  - We can include an interactive platform for communication between the Portal developers and teachers, as well as teachers. One idea is to create a page for benchmarking criteria, asking questions, and sharing experiences and information, accessible only to registered teachers. This would allow for open discussions among teachers in a closed forum, similar to TEAL's (2024b) discussion forum.

Conclusion

We developed a Speaking Assessment (SA) Portal to address the needs of Japanese senior high school English teachers for online SA resources that can be used for teacher training and self-study. The Portal includes various videos, each consisting of a task, a rubric, rubric-based scores, a transcription, an explanation of the scores, and a worksheet in the "SA examples and explanations" section. Other sections include "Tips for conducting speaking tests" and "Useful websites and resources."

We then examined the usability of this Portal by using feedback from teachers who accessed the website to assess its quality and identify areas for improvement. The first research question explored how Japanese teachers of English perceive the Portal's usefulness after using it. The responses in Phase 1 indicated that most teachers viewed the Portal positively. In particular, "SA examples and explanations" was considered the most interesting and appropriate by the intended users.

The second research question investigated what areas of the Portal these teachers identified for improvement. Numerous suggestions were made, ranging from adding more information, materials,

examples, and visual aids, to including more diverse types of task formats and rubrics. We have addressed most of these suggestions from the participating teachers and plan to further develop the content and functionality of this Portal website.

There are limitations in the current preliminary usability study. First, we gathered feedback from a relatively small number of teachers. Involving a more diverse group of participants could provide a wider range of perspectives useful for revision. Second, we did not employ extensive questionnaires or interviews to collect data on teacher perceptions. A more comprehensive approach, using a mixed-methods research design to gather teachers' perceptions from various viewpoints, would provide more detailed insights and help improve the Portal (see Shen et al., 2015, for a method example).

Regarding the practical implications derived from the current study, the Portal can be useful for teacher training and independent study, as indicated by teachers' perceptions. Moreover, asking intended resource users to provide feedback through various open-ended questions is critical. To address teachers' interests and concerns, content developers should involve users from diverse backgrounds (e.g., current and former teachers, teacher trainers, teachers with and without knowledge of language assessment and speaking assessment, as was done in the current study). To effectively utilize the feedback from users, content developers need to plan ahead and allocate sufficient time to receive comments and revising resources at multiple stages of development. These efforts would facilitate communication between content providers and users, ultimately benefiting the dissemination of content—in this case, SA principles and practices.

Acknowledgement

Notes

1. The following is the description from the FAQ section of the Portal website.

- How can we assess speaking interaction (dialogues) and presentation (monologues) efficiently and fairly when there are approximately 40 students per class, with five classes in one school per year?

    - It is important to create a yearlong plan and decide when to conduct speaking tests in relation to teaching. Tests should be planned to focus on validity while also considering reliability and practicality. When considering practicality, the following questions arise: (a) How many lessons can be used for a class with 40 students? (b) Can tests be scored outside of class time? and (c) How many minutes per student can be allocated for conducting speaking tests?

        Based on answers to these questions, it is possible to select one out of four patterns, as described in "Administration of speaking tests" ((v) in-class administration and in-class scoring, (x) in-class administration and out-of-class scoring, (y) out-of-class administration and scoring on the spot, and (z) out-of-class administration and scoring after the test; see Koizumi, 2022b, p. 154). Using the selected option, we can concretely decide on a test format (e.g., teacher-led interviews, pair conversations, group discussions to measure oral interaction) and a rubric. We will also decide

whether to focus on presentation or interaction, as well as what specific abilities we would like to measure while considering teaching objectives and what activities are conducted in lessons.

For example, when we can only use (a) one 50-minute lesson (b) with tests scored within the lesson, we can use only 40 minutes for a speaking test because 10 minutes are needed for explanation. Then, we can use (c) one minute per student (maximum of 40 seconds for speaking time). Test formats that align with this requirement include speech to measure monologues, and group discussions to measure interactions. In Matsuo's practice (Matsuo, 2019), all 40 group discussions are tested and scored within a 50-minute lesson.

Regarding the number of speaking tests to be conducted in a year, some schools have one per term with three speaking tests in total in a year. Others have speaking tests around the time of the term tests, so they have five speaking tests in total per year.

Koizumi (2022a) shows speaking test samples conducted in the second year at a school and discusses how teachers maintain validity and reliability.

2. The following is the description from in the FAQ section of the Portal website.

- Could you please improve the situation where sample speeches are difficult to hear in some videos due to recording conditions and students' voice volumes?

    - Some videos were difficult to hear because actual test videos were used on the Portal. However, in real-test scoring, videos that do not have ideal conditions are still scored, so this difficulty may reflect real-life situations. Although we have recordings made with voice recorders and could overlay those sound files onto the videos, we chose not to do so and kept the original video sounds. This is because sounds that are undetectable by the ears may be picked up by voice recorders, which can differ substantially from what is heard during speaking tests.

3. Teacher: I provided an example of the club activity task of the SA Portal website.

- Current description:

    - Role play task: Talking about club activities and hobbies

    - Setting:

        - Teacher: An international student in the same class. The student wants to join a club activity, so they ask questions.

        - Student: A student who wants to make friends with the international student. Answers questions and asks questions.

In addition, more can be added to the situation related to an international student, as follows:

- Additional setting (see italics for the addition):

    - Teacher: An international student in the same class. The student wants to join a club activity, so they ask questions. They want to experience Japanese culture (or a sport specific to Japan) in a club activity. They do not have much money and would like to join a club without the need to buy tools for the club.

References

Assessment and Evaluation Language Resource Center (2024). *Resources*. https://aelrc.georgetown.edu/resources/

British Council. (2024a). *Doga de miru eigo yongino hyoka no pointo* [Tips for assessing English four skills through videos]. https://www.britishcouncil.org/exam/english/aptis/research/assessment-literacy

British Council. (2024b). *How language assessment works.* https://www.britishcouncil.org/exam/english/aptis/research/assessment-literacy

British Council. (n.d). *Language assessment resources*. https://www.teachingenglish.org.uk/professional-development/teachers/assessing-learning/articles/language-assessment-resources

Fulcher, G. (2024, September 11). *Language testing resources website*. https://languagetesting.info/

Indiana University (2024). *TBLT language learning task bank*. https://tblt.indiana.edu/tasks/index.html

Japan Language Testing Association. (n.d.). *Workshop video/web tutorial*. https://jlta2016.sakura.ne.jp/?page_id=808

Kaneko, J. (2019). *Chuko renkei wo fumaeta eigo jugyo niokeru activity to performance test kaihatsu nikansuru chosakenkyu* [Survey research into activities and performance test development for English lessons based on collaborations between junior and senior high schools]. 2018 report of a survey research project on teacher pre-service training. https://www.gakushubunka.jp/scholarship/kenkyugaiyou_kaneko_h30.pdf

Koizumi, R. (Ed.). (2022a). *Jitsurei de wakaru eigo speaking test sakusei gaido* [A practical guide for developing English speaking tests]. Taishukan Publishing.

Koizumi, R. (2022b). L2 speaking assessment in secondary school classrooms in Japan. *Language Assessment Quarterly*, *19*(2), 142–161. https://doi.org/10.1080/15434303.2021.2023542

Matsuo, M. (2019). *Jirei hokoku: Tesuto ga totatsumokuhyo to shido niataeru eikyo: Semina repoto* [Case study: Effects of tests on course goals and teaching: Seminar report]. British Council. https://www.britishcouncil.jp/programmes/english-education/japan/report/assessment2018-seminar/case1

MEXT (Ministry of Education, Culture, Sports, Science and Technology). (2022). *Shido to hyoka no ittaika nimuketa koko gaikokugoka niokeru pafomansu tesuto sanko shiryo (shidosya yo shiryo)* [Reference guides of performance tests in teaching foreign languages at upper senior high schools: Toward the integration of teaching and assessment: Materials for instructors]. https://www.mext.go.jp/content/20220705-mxt_kyoiku01-1000021347-1.pdf

Nakatsuhara, F., May, L., Lam, D., & Galaczi, E. (2018). *Learning oriented feedback and interactional competence* (*Research Notes*, Vol. 70). Cambridge Assessment English. https://www.cambridge-exams.ch/research-notes-issue-70-learning-oriented-feedback-and-interactional-competence

National Institute for Educational Policy Research. (2023). *Shido shiryo jirei shu* [Handbook of teaching materials and case examples]. https://www.nier.go.jp/kaihatsu/shidousiryou.html

Poehner, M. E., & Inbar-Lourie, O. (Eds.). (2020). *Toward a reconceptualization of second language classroom assessment: Praxis and researcher-teacher partnership*. Springer.

Shen, H., Yuan, Y., & Ewing. R. (2015). English learning websites and digital resources from the perspective of Chinese university EFL practitioners. *ReCALL*, *27*(2), 156–176. https://doi.org/10.1017/S0958344014000263

Tamura, T. (2022). *Korede wakaru, dekiru! Shogakko gaikokugo pafomansu tesuto: Hanasukoto yaritori no hyoka* [Understand and master it! Performance tests for a foreign language English in elementary schools: Assessing speaking Interaction; DVD].

Tando, H. (2023). The investigation of issues of speaking performance evaluation in elementary schools in Aomori Prefecture. *TELES (Tohoku English Language Education Society) Journal*, *43*, 84–96. https://doi.org/10.57539/telesjournal.43.0_84

TEAL (Tools to Enhance Assessment Literacy). (2024a). *Common oral assessment tool*. https://teal.global2.vic.edu.au/assessment-tools/common-oral-assessment-tasks/

TEAL. (2024b). *Discussion forum*. https://teal.global2.vic.edu.au/discussion-forum/

TEAL. (2024c). *Tools to Enhance Assessment Literacy for teachers of English as an additional language*. https://teal.global2.vic.edu.au/

Tono, Y. (2022). *CEFR-J CAN-DO tesuto: Sampuru bajon* [CEFR-J CAN-DO test: Sample version] (Version 1.0). Retrieved March 17, 2024, from https://www.cefr-j.org/download.html#cefrj_testasks

Tono, Y., & Negishi, M. (Eds.). (2020). *Kyozai tesuto sakusei notameno CEFR-J risosubukku* [The CEFR-J resource book: Reference level descriptions and test development]. Taishukan Publishing.

Appendix A

Questionnaire 1 in Japanese

Note. This was answered after reviewing the overall site.

1. お名前をお願いします。

2. Website の中で、最初にパッと見てみて、面白い、読んでみたいと思ったものを選んでください（複数回答可）。

   A. スピーキングテストのコツ
   B. スピーキングテストの実例と解説
   C. 役立つ Website・参考資料
   D. プロジェクトメンバー紹介

3. 「スピーキングテストのコツ」について感想をお願いします。（例：だいたい内容は知っていた。〜についてさらに知りたい。レイアウトは〜だ）

4. （「スピーキングテストの実例と解説」については、後で詳細に見ていただきますので、飛ばします。）「役立つ Website・参考資料」について感想をお願いします。（例：だいたい内容は知っていた。〜についてさらに知りたい。レイアウトは〜だ）

5. 「プロジェクトメンバー紹介」について感想をお願いします。（例：だいたい内容は知っていた。〜についてさらに知りたい。レイアウトは〜だ）

6. 本 Website を、高校の先生方などに使っていただくために、何かあったらよいと思う内容や機能はありますか？あれば書いてください。

7. 他に何か感想かご意見があればよろしくお願いいたします。

Appendix B

Questionnaire 1 in English

Note. This was answered after reviewing the overall site. It was translated into English by the first author.

1. Please write your name.

2. Please select all the items that you found interesting or would like to read. (Multiple answers were allowed.)

   A. Tips for conducting speaking tests

   B. SA examples and explanations

   C. Useful websites and resources

   D. Project members

3. Please write your impression about "A. Tips for conducting speaking tests" (e.g., "I knew almost all the content"; "I want to know more about …"; "The layout is ….")

4. (Please skip "B. SA examples and explanations." You will be asked to read it later.) Please write your impression about "C. Useful websites and resources." (e.g., "I knew almost all the content"; "I want to know more about …"; "The layout is ….")

5. Please write your impression about "D. Project members." (e.g., "I knew almost all the content"; "I want to know more about …"; "The layout is ….")

6. Please write any content or functions, if any, that this Portal should have to facilitate the use from senior high school teachers and others.

7. Please write any other opinions.

Appendix C

Questionnaire 2 in Japanese

Note. This was answered after reviewing "B. SA examples and explanations."

1. お名前をお願いします。

2. 「Speaking test の実例と解説」の中でご覧になったタスクを選んでください（複数回答可）。

   A. 教員と Role play 映画 タスク 1：映画に誘う (CEFR-J A1.3)

   B. 教員と Role play 道案内 タスク 2：道案内をする (CEFR-J A2.1)

   C. 教員と Role play 学校 タスク 3：学校を紹介する (CEFR-J B1.1)

   D. 教員と Role play 教育 タスク 3：子どもの教育の改善を提案する (CEFR-J B2.1)

   E. 話すこと（発表・スピーチ）・話すこと（やり取り・ペアでの質疑応答）

3. 「Speaking test の実例と解説」について全体的な感想をお願いします。（例：だいたい内容は知っていた。〜についてさらに知りたい。レイアウトは〜だ）

4. 「Speaking test の実例と解説」の中の細かな点について気になった点等書いてください。

5. 「Speaking test の実例と解説」を、高校の先生方などに使っていただく（例：自己研究、校内研修、スピーキングテスト実施前の打ち合わせ）ために、何かあったらよいと思う内容や機能はありますか？あれば書いてください。

6. 他に何か感想かご意見があればよろしくお願いいたします。

7. 今回の Website 確認にかけてくださった時間はどのくらいですか？（例：〜分、〜時間）

（謝礼に関する質問は、省略）

Appendix D

Questionnaire 2 in English

Note. This was answered after reviewing "B. SA examples and explanations." It was translated into English by the first author.

1. Please write your name.

2. Please select all the tasks that you saw in the "B. SA examples and explanations." (Multiple answers were allowed.)

   Task 1: Inviting your friend to see a movie (CEFR-J A1.3)

   Task 2: Showing the way (CEFR-J A2.1)

   Task 3: Introducing your school (CEFR-J B1.1)

   Task 4: Suggesting a way to improve child education (CEFR-J B2.1)

   Task 5: Making a speech and asking questions and answer them in a pair

3. Please write your overall impression about "B. SA examples and explanations." (e.g., "I knew almost all the content"; "I want to know more about …"; "The layout is ….")

4. Please write any points in detail to improve the site in "B. SA examples and explanations."

5. Please write any content or functions, if any, that this Portal should have to facilitate the use from senior high school teachers and others (e.g., self-study, within-school training, meeting before the administration of a speaking test).

6. Please write any other opinions.

7. How long did you spend reading the Portal and writing your opinions? (e.g., … minutes, … hours)

(Other questions related to honorariums were omitted here.)